ERDC/CERL CR-09-1

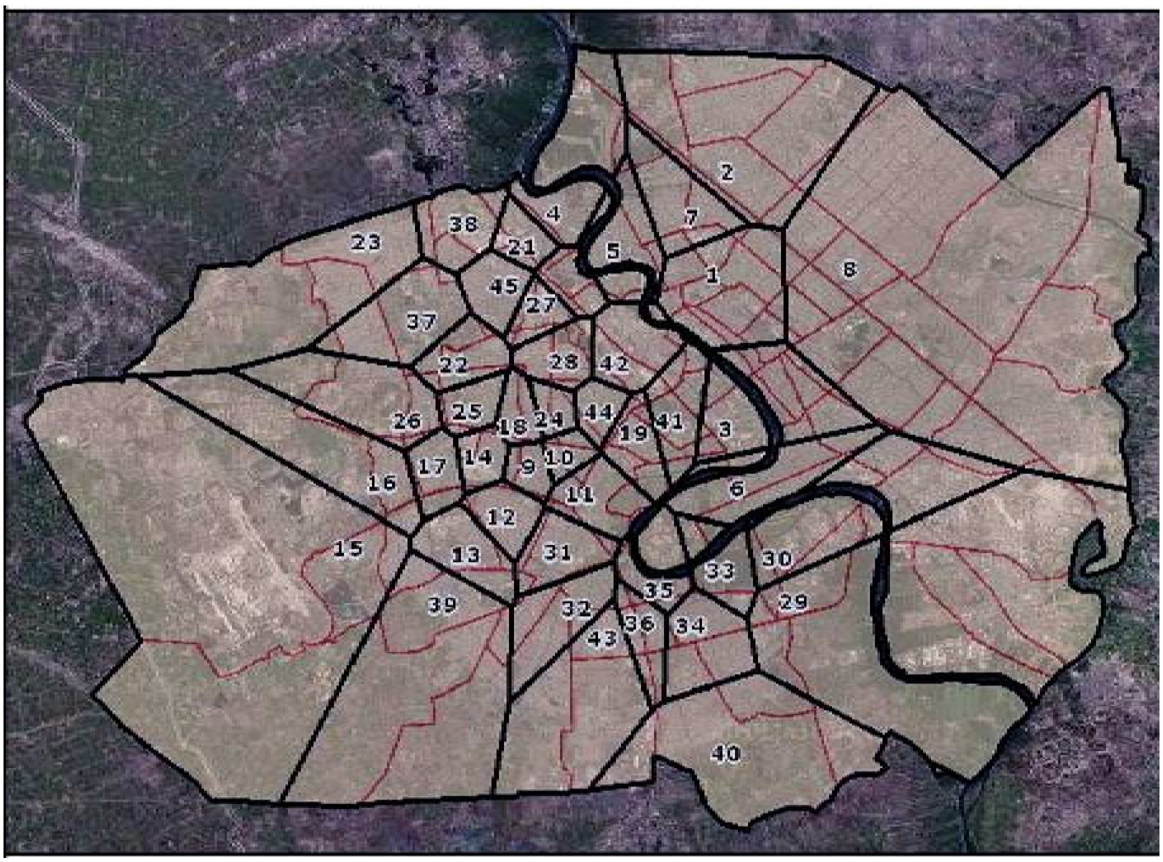
Construction Engineering Research Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

Approaches Towards the Identification of Patterns in Violent Events, Baghdad, Iraq

Luc Anselin and Gianfranco Piras

May 2009



Approved for public release; distribution is unlimited.

ERDC/CERL CR-09-1
May 2009

Approaches Towards the Identification of Patterns in Violent Events, Baghdad, Iraq

Luc Anselin and Gianfranco Piras

GeoDa Center, School of Geographical Sciences
Arizona State University
Tempe, AZ 85287-5302
http://geodacenter.asu.edu/

Final Report

Approved for public release; distribution is unlimited.

Prepared for Headquarters, U.S. Army Corps of Engineers
Washington, DC 20314-1000

Abstract: This work analyzed the spatial distribution of violent events as constructed from a content analysis of open source news reports. Data on 112 variables was available for 45 neighborhoods. The small sample limited the analysis to those 32 variables with at least four observations. The statistical analysis was done both for the original measures of event counts by neighborhood, and for binary variables that indicated the presence of events. Test statistics for spatial autocorrelation were computed for global patterns and local patterns, including global and local Moran's I, Geary's c, Moran Scatterplot, join count statistics, and local join counts.

There was little evidence of systematic spatial structure at the neighborhood scale. Only for a variable indicating internal between hayy migration was there consistent indication of positive spatial autocorrelation, or clustering. Several other variables showed significant negative spatial autocorrelation at the local scale, suggesting that neighborhoods where violent events occurred are surrounded by neighborhoods without violent events. A few neighborhoods were consistently identified as the locus of a spatial outlier, suggesting some patterning. A finer spatial scale might reveal more complex spatial patterns. The current data do not allow this to be investigated.

Approaches Towards the Identification of Patterns in Violent Events, Baghdad, Iraq

Luc Anselin
Gianfranco Piras

GeoDa Center, School of Geographical Sciences
Arizona State University
Tempe, AZ 85287-5302

http://geodacenter.asu.edu/

February 19, 2009

Report in Partial Fulfillment of Cooperative Agreement W9132T-08-2-005

Contents

List of Tables

List of Figures

Acknowledgments

The research reported on in this report was supported by Cooperative Agreement W9132T-08-2-005 between U.S. Army ERDC-CERL and the GeoDa Center for Geospatial Analysis and Computation at Arizona State University. Many thanks to the CERL team, and especially William Meyer, Lucy Whalley and David Krooks for providing the data and advice regarding the subtleties of Iraqi sources, place names and content analysis. Charles Schmidt of the GeoDa Center team provided assistance with data management and GIS operations. Julia Koschinsky provided logistical support. The report and its findings are those of the authors alone, and should not be attributed to ERDC/CERL or the U.S. Army.

Executive Summary

In this report, we analyze the spatial distribution of violent events as constructed from a content analysis of open source news reports. Data on 112 variables was available for 45 neighborhoods. The sparsity of the sample limited the analysis to those 32 variables for which there were at least four observations. The statistical analysis was carried out both for the original measures of counts of events by neighborhood, as well as for binary (0-1) variables that only indicated the presence or absence of events.

Test statistics for spatial autocorrelation were computed for global patterns and local patterns. These include global and local Moran's I, Geary's c, Moran Scatterplot, join count statistics and local join counts.

Overall, there was little evidence of systematic spatial structure at the scale of the neighborhood. Only for a variable indicating internal between hayy migration was there consistent indication of positive spatial autocorrelation, or clustering. Several other variables showed significant negative spatial autocorrelation (spatial outliers) at the local scale, suggesting that neighborhoods within which violent events occurred are surrounded by neighborhoods without violent events. A few neighborhoods were consistently identified as the locus of a spatial outlier. This suggests some patterning, but more extensive information is needed to characterize this more accurately.

Finally, it should be kept in mind that a finer spatial scale might reveal more complex spatial patterns. The current data do not allow this to be investigated.

1 Introduction

This report provides the methodological context and application of various techniques to assess patterns in violent events in Baghdad, Iraq. The original motivation was to assess the extent to which data culled by means of content analysis from open source news reports could be used to construct quantitative measures of the presence of significant patterns, such as hot spots and spatial outliers. In the process, several data quality issues needed to be addressed. The results presented here are only preliminary, in the sense that ultimately insufficient data of acceptable quality could be extracted from the original data sources to allow meaningful spatial statistical inference. The evidence on spatial patterns in Baghdad is therefore limited, although approaches are outlined that could be followed in a situation where more extensive data are available (e.g., using classified sources of information).

The report is organized in six remaining sections. First, a brief overview is given of the fundamental methodological issues and the statistical tests most commonly used in practice are presented. This also includes a discussion of new techniques proposed to deal with binary (zero-one) data. Next, the data and software are addressed, before moving on to the actual empirical analysis. The latter is carried out using both a global and a local perspective for the original counts of events and for a binary indicator of presence/absence of the events. We close with some concluding remarks.

2 Methodological Issues

In this section, we review some relevant methodological issues pertaining to the analysis of spatial data. In many instances empirical data not only contain information on the variable (or attribute) of interest but also on the geographic location where the particular value was observed. Spatial methods are required when the structure of the data violates standard statistical assumptions, such as independence. For example, the so-called first law of geography states that observations close in space tend to be more similar. This is also referred to as positive spatial autocorrelation, a crucial concept in the analysis that follows.

We start with a review of types of spatial data in Section 2.1. Next, we elaborate on the concept of spatial autocorrelation, providing a formal definition in Section 2.2 as well as highlighting some of its special characteristics. Section 2.3 deals with one of the main problems when applying spatial analysis, the definition of similarity and neighboring relationships. Finally, Section 2.4 provides a brief discussion of statistical inference in the context of spatial autocorrelation measures.

2.1 Types of Spatial Data

The classification we adopt follows Cressie (1993) and characterizes spatial data by the nature of the spatial domain (D). The spatial domain is continuous when a variable Z can be observed everywhere within the domain. In other words, between any given pair of locations one can theoretically place an additional infinite number of observations. On the other hand, a spatial domain is discrete when the number of locations can be enumerated, even though their actual number can

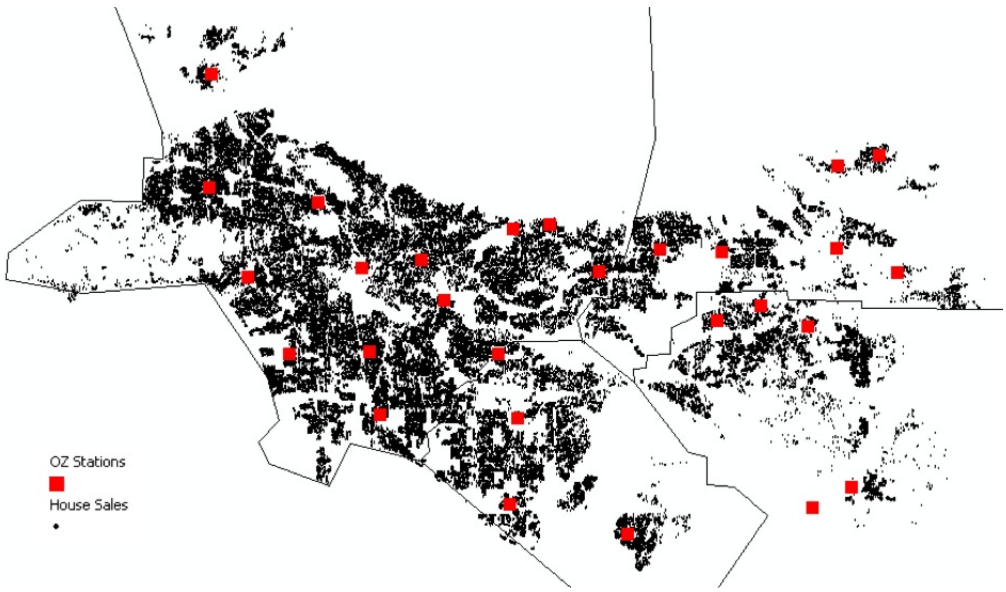


Figure 1: Geostatistical Data, House Sales and Ozone Stations, Los Angeles, CA

approach infinity. An example of this situation is when administrative units, such as districts or provinces are considered. In the literature and in practice, the types of data that correspond to these two situations are referred to as geostatistical data and lattice data, respectively.

Geostatistical data corresponds to a continuous and non-stochastic domain. As an example, consider measuring air pollution. Air pollution could be recorded at any location, continuously across space. In practice, however, the level of pollution is recorded at a finite number of specific monitoring locations (see Figure 1). In contrast, for lattice (or regional) data the domain is non-stochastic but discrete. A typical example of lattice data are the U.S. counties, or the neighborhoods of a city, or, more in general, data sets that provide information about every site in the domain (see Figure 2).

Both geostatistical and lattice data share the common characteristic of a non-

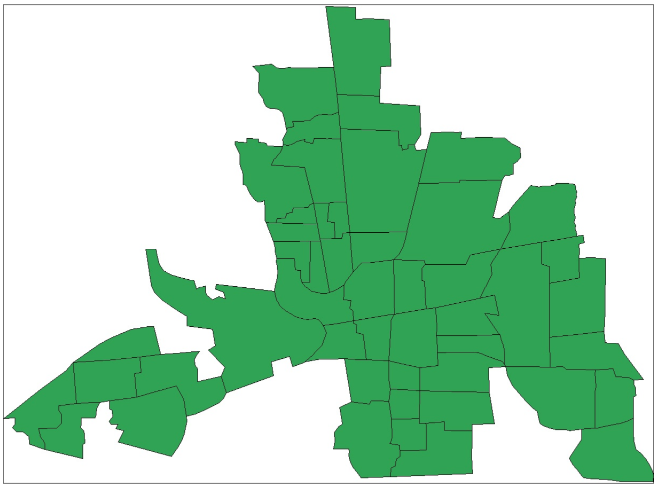


Figure 2: Lattice Data, Neighborhoods in Columbus, OH

stochastic domain. In contrast to this, so-called point pattern analysis deals with the spatial distribution of events across space. In other words, the locations of the events are stochastic and thus the domain (possible locations) is random. Formally, the collection of points corresponding to a random set D is termed a point pattern (see Figure 3).

2.2 Spatial Autocorrelation

Generally speaking, autocorrelation is the correlation of a variable with itself. In statistics, as well as in mainstream econometrics, a huge emphasis has been placed on the presence of serial correlation over time. In time series analysis, past values determine the present value of a variable. However, clearly, the present does not determine the past. In other words, the structure of the dependence and autocorrelation is unidirectional. The definition of dependence in space is not as straightforward as in the time domain. Dependence in space is multi-directional

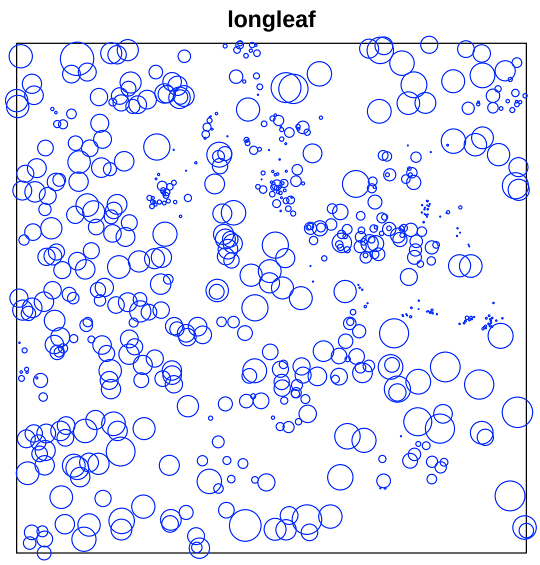
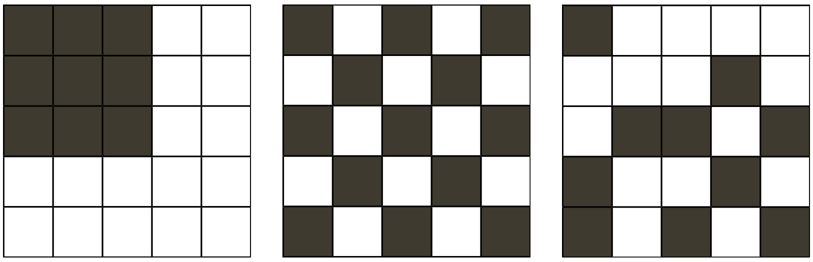


Figure 3: Point Pattern Data, Locations of Trees

in that what happens at one site may influence all other sites, but also may be influenced by every other location. Therefore, if the variable under observation is distributed across space, the term autocorrelation will refer to the correlation between the value of the same variable at different locations.

Spatial autocorrelation can then be defined as the coincidence of value similarity with locational similarity (Anselin and Bera 1998). Two types of spatial autocorrelation occur, characterized as positive or negative spatial autocorrelation (see Figure 4). If there is positive spatial autocorrelation, then observations with high (or low) values of a variable tend to cluster in space (neighbors, or locations with similar locations have similar values). In other words, proximity in space corresponds with attribute similarity (panel (a) of Figure 4). Conversely, in the presence of negative spatial autocorrelation, locations tend to be surrounded by neighbors having dissimilar values. Negative spatial autocorrelation implies a checkerboard pattern of values, as shown in panel (b) of Figure 4. Positive spatial autocorrelation is by far more intuitive with a clear interpretation, and tends to be more commonly the focus of attention. Negative spatial autocorrelation is interesting in that it points to spatial heterogeneity, or the dissimilarity of locations.

It should be noted that the point of departure in spatial statistical analysis is the absence of any spatial structure, be it positive or negative. This is referred to as complete spatial randomness (see Figure 4 panel (c) for an example). Spatial randomness implies that there is no relation between the location of observations and their values. As pointed out in Cressie (1993), the consequence of positive spatial autocorrelation in data is a decrease of information with respect to the uncorrelated (spatially random) counterpart. Proper statistical inference should explicitly account for this loss of information in estimation and hypothesis testing.



(a) (b) (c)

Figure 4: Examples of Spatial Autocorrelation Over a 5×5 Regular Grid. Positive Spatial Autocorrelation (Panel a), Negative Spatial Autocorrelation (Panel b) and Spatial Randomness (Panel c).

2.3 Spatial Weights

In defining spatial autocorrelation, the formal expression of the “proximity” between spatial observations is a crucial concept. Many different criteria can be used to define “neighbors.”

The most intuitive way to proceed would be to estimate the full variance-covariance matrix (or correlation) directly from the observed data. However, for a set of n observations there will potentially be $n \times n$ covariance terms. This is an example of the incidental parameter problem, a lack of degrees of freedom in the data which makes estimating these parameters impossible in practice. The solution to this problem consists of imposing some constraints in order to reduce the number of parameters to be estimated. The tool that is widely used to represent the spatial structure of a set of locations is the so-called spatial weights matrix.

A spatial weights matrix (W) is an $n \times n$ positive and symmetric matrix that defines the structure of the interaction between each pair of spatial units. Each observation appears both as a row and as a column in the matrix. In each row, the non-zero elements of the matrix express for a given observation which other obser-

vations are considered to be its neighbors. Each non-zero element (w_{ij}) indicates the intensity of the relationship between the corresponding cross sectional units (i and j). By convention, the diagonal elements w_{ii} are set to zero, which excludes so-called self-similiarity, or the possibility that an observation is considered to be its own neighbor. The weights matrix is typically used in row standardized form (so that all its elements range between 0 and 1). To obtain the row standardization, each element is divided by the sum of all the row elements. An important consequence of the row standardization is that the resulting matrix will not necessarily be symmetric, even though the original definition of neighbors is symmetric.

A crucial aspect in the construction of the spatial weights matrix is to determine which elements should be different from zero. There is little guidance on this from either a theoretical or practical perspective. Typically, the description of the neighborhood set relies on geographical criteria. The most straightforward case is when two observations are considered neighbors if they share a common edge. When this is the case, the corresponding element of W will be set to one (binary contiguity). On a regular grid, different definitions of binary contiguity can be applied. In analogy with the game of chess, they are referred to as rook (share a common edge), bishop (share a common vertex), or queen (both edge and vertex). In practice, rook and queen are the most commonly used criteria.

An alternative to the contiguity criterion is to use distance between observations as the basis for the construction of weights. Non-zero elements of W correspond to pairs of observations within a critical distance of each other (i.e., $w_{ij} = 1$ if $d_{ij} < d$, where d_{ij} is the physical distance between observations i and j, and d is a distance cut-off value). Sometimes a function (inverse or squared inverse) of the distance is used directly as input in the spatial weights matrix. In the literature,

this is referred to as direct representation.

A further example of geographical weights are the so-called k-nearest neighbors. Such neighbors are based on the sorted distance between a unit and all other units. The order k determines how far the neighbor set reaches. Note that the actual magnitude of the distances between units is not important, it is the relative ranking of distances that determines the neighbor set for each location. This criterion is not symmetric, which complicates the computation of test statistics and model parameters in some instances.

Geographical definitions of the neighbor set are by far the most widely used in practice. However, many other approaches have been proposed in the literature as well (see Case 1991, 1993, Conley and Ligon 2002, Conley and Topa 2002, Case et al. 1993, among others), including semiparametric weights (Pinkse et al. 2002).

2.4 Inferential Approaches

Inference about spatial autocorrelation is based on a hypothesis test of the null hypothesis that the data are spatially random. In different contexts (geostatistical data, lattice data, point patterns), this takes on a slightly different form, but the essence is that, under the null, space does not matter.

The general idea is to assess whether the magnitude of the observed test statistics computed from the data is unusual under the assumption of spatial randomness. If the value of the statistic is particularly extreme, the null hypothesis should be rejected and evidence is found of the presence of a spatial structure. As discussed in Schabenberger and Gotway (2005) there are four general approaches to assess the significance of a test statistic for spatial autocorrelation.

A permutation test is based on the idea that spatial randomness can be operationalized in the form of a random assignment of each observed value of the variable to each spatial unit with equal probability. The total such possible arrangements for n observations equals the number of possible permutations ($n!$). Any spatial autocorrelation statistic is then computed for each of these random spatial arrangements to yield a reference distribution. The observed value of the test statistic (i.e., the value computed from the actual data) is compared to the reference distribution to assess pseudo-significance. Typically, if the observed value is in the tail of the reference distribution, the null hypothesis of spatial randomness is rejected.

The main limitation of the permutation approach is that even for small n the number of arrangements will quickly become very large and create impractical computational demands. Therefore, in practice, rather than considering the exhaustive set of permutations, a Monte Carlo approach is used instead. This creates a subset of all possible arrangements by randomly reshuffling the data a predetermined number of times. The number of replications determines the smallest pseudo-p value that can be obtained. For 99 permutations, this is $p = 0.01$, for 999 permutations, $p = 0.001$.

In contrast to this computational (non-parametric) approach, a parametric perspective starts from strong assumptions about the underlying distribution of the variable. Under suitable regularity conditions, this allows for the computation of the moments of the test statistic under the null hypothesis, and sometimes the (asymptotic) distribution of the statistic as well. The most commonly used approach is to assume a Gaussian distribution for the variable, which in most cases leads to a normal approximation of the distribution of the test statistic. In practice,

the mean and variance of the test statistics (under the null hypothesis) are used to construct a standardized z-value, which is then employed to assess significance.

As an alternative to the Gaussian assumption, equal probability (randomization) is commonly used as well. Again, under a set of suitable regularity conditions, in most instances this results in an asymptotic normal approximation of the distribution of the test statistic.

3 Statistics for Spatial Autocorrelation

Probably the most widely used statistic for spatial autocorrelation is the so-called Moran's *I* test (originally presented in Moran 1950). We consider this in Section 3.1 for both global and local versions of the test statistic. Section 3.2 deals with Geary's *c* statistic (Geary 1954), again both global and local. Finally, in Section 3.3, we turn to global and local measures of spatial autocorrelation for discrete (binary) data, based on the principle of join counts.

3.1 Moran's *I*

The formal expression for Moran's I statistic for variable x is:

$$I = \frac{n}{S_0} \sum_i \sum_j w_{ij}(x_i - \mu)(x_j - \mu) / \sum_i (x_i - \mu)^2 \tag{1}$$

where n is the total number of observations, $S_0 = \sum_i \sum_j w_{ij}$ is the sum of all the elements of the spatial weight matrix, and μ is the mean of x over all the observations. Note that if W is row-standardized, then $S_0 = n$ and the ratio n/S_0 is equal to one.

Moran's I resembles a correlation statistic since the numerator is a measure of the covariance between the variable at different locations, and the denominator has the typical aspect of a measure of the variance. The distribution of the test statistic under the null is asymptotically normal. The moments of Moran's I can be derived analytically, using either a normality assumption (of the underlying variable) or equal probability (randomization). In both instances, the expected value of the statistic turns out to be $E(I) = -1/(n-1)$ (Cliff and Ord 1981). However, the variance differs depending on the assumption used. Under the assumption of normality, the second moment is (see Cliff and Ord 1973, Ch1, p. 15):

$$E_N(I^2) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{S_0^2(n^2-1)} \tag{2}$$

while under the assumption of randomization it is:

$$E_R(I^2) = \frac{n[(n^2-3n+3)S_1 - nS_2 + 3S_0^2] - b_2[(n^2-n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} \tag{3}$$

where n is again the total number of observations, $S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij}+w_{ji})^2$, $S_2 = \sum_{i=1}^{n}\left[\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right]^2$ and S_0 has been defined above. Also, b_2 is the sample kurtosis coefficient, m_4/m_2^2, where m_j is the $j-th$ sample moment of x around the sample mean. Inference is generally based on a standardized z-value computed with either one of these moment expressions.

The interpretation of the index is as follows. When $I > E(I)$, then a spatial unit tends to be surrounded by locations with similar attributes. In other words, there is spatial clustering of either low/low or high/high values. The strength of positive spatial autocorrelation tends to increase with $[I - E(I)]$. On the other hand, when

$I < E(I)$, an observation will tend to be surrounded by neighbors with dissimilar values, suggesting negative spatial autocorrelation.

As such, the magnitude of the statistic is not sufficient to conclude significance. Irrespective of the magnitude of the statistic relative to its mean, there is only evidence of departure from the null hypothesis of spatial randomness when the statistic is large enough to be deemed significant.

3.1.1 Local Moran

Local Indicators of Spatial Association (LISA), according to the definition in Anselin (1995), are statistics that satisfy two requirements:

1. For each location, provide an indication of the extent to which there is significant clustering of similar values (local clusters) or dissimilar values (spatial outliers) around that location.

2. The sum of the local statistics over all observations is proportional to a global indicator of spatial association.

In contrast to global measures, local measures of spatial association examine spatial dependence for subsets of the data. Thus, while a global measure yields a single value for the entire data set, local measures result in as many statistics as there are spatial units in the sample. Getis and Ord (1996) list a number of advantages of local statistics, including:

1. Identification of hot and cold spots.

2. Identification of the scale at which there is no distinguishable association of data values.

3. Verification of stationarity for a given study regions

As shown in Anselin (1995), a local version of Moran's I can de expressed as:

$$I_i = (x_i - \mu) \sum_{j \in J_i} w_{ij}(x_j - \mu), \tag{4}$$

where the variables have the same meaning as in (1) and J_i is the set of neighbors of observation i. Anselin (1995) also derives the expression for expected value and variance under a randomization assumption. These moments can be used for statistical inference similar to the approach used for the global Moran. However, the asymptotic approximation of the distribution of the local Moran under the null based on these theoretical moments is rather poor. A preferred approach to inference is conditional randomization, in which a series of permutations is carried out for each observation in turn (see Anselin 1995, for details).

3.1.2 Moran Scatterplot

When the weights matrix W is row-standardized, an interesting interpretation of Moran's I statistic follows as the slope in a regression line. As a matrix expression, the numerator of Moran' I can be written as $z'Wz$, where $z_i = (x_i - \mu)$. Using the same notation, the denominator can be written as $z'z$. As a result, the statistic becomes $I = z'Wz/zz$. This corresponds to the slope in a linear regression of Wz (dependent variable) on z (explanatory variable).

A graphical depiction of this result is obtained in a so-called Moran scatterplot (Anselin 1996), in which the vertical axis represents the spatially lagged variable Wz, and the horizontal axis the original variable z. This is the opposite of what a spatial autoregressive regression would suggest, but it is the appropriate arrange-

ment to find Moran's I as the slope of the linear regression through the scatterplot cloud.

A Moran scatterplot is a convenient tool to identify potential clusters and spatial outliers (Anselin 1996). It provides a visual exploration of spatial autocorrelation. If most of the observations fall either in the upper right or lower left quadrants, there is evidence of positive spatial autocorrelation. Conversely, if the observations are mostly concentrated in the lower right or upper left quadrants, this suggests negative spatial autocorrelation. The four quadrants in the scatterplot provide the basis to classify observations into four types of spatial autocorrelation. Spatial clusters occur in the form of positive high-high associations in the upper right quadrant, and positive low-low associations in the lower left quadrant. Spatial outliers are present in the form of negative high-low associations in the lower right quadrant, and negative low-high associations in the upper left quadrant. Note that this only suggests the type of association, but does not establish significance. A formal permutation test would be necessary to establish the latter.

A Moran scatterplot is illustrated in Figure 5.

3.2 Geary's c

Geary's c statistic can be expressed as:

$$c = \frac{(N-1)}{2S_0} \sum_i \sum_j w_{ij}(x_i - x_j)^2 / \sum_i (x_i - \mu)^2, \tag{5}$$

using the same notation as in the expression for the Moran's I. Similar to the procedures outlined in Section 3.1, inference can be based either on a normality assumption or on a randomization approach. Interestingly, the expected value

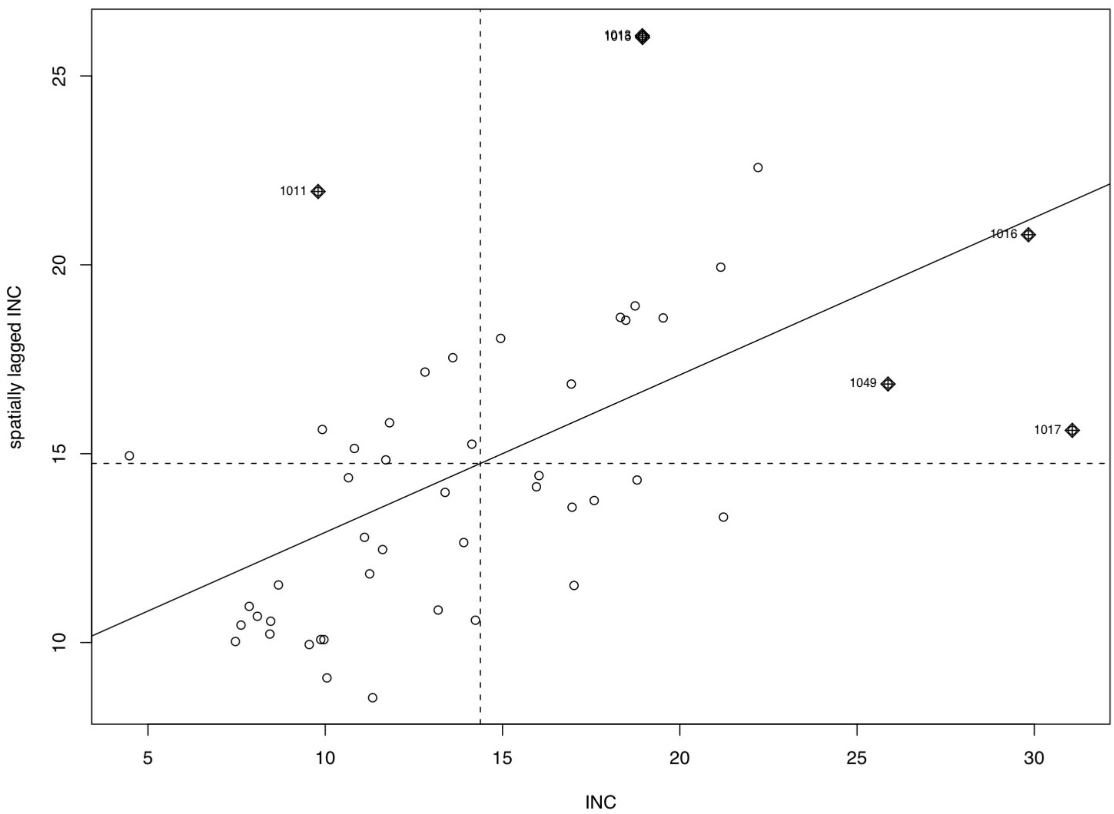


Figure 5: Moran Scatterplot. Variable Income, Columbus Dataset.

of the statistic under the null is $E(c) = 1$ in both approaches. Consequently, it does not depend on which variable is being considered, on the spatial structure implied by W, or on the sample size. However, the variance differs between the two assumptions. Under the assumption of normality, the variance is:

$$\mathrm{Var_N}(c) = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0^2}, \tag{6}$$

whereas under the randomization assumption, it becomes (see Cliff and Ord 1973, for technical details):

$$\begin{aligned} \mathrm{Var_R}(c) \quad = \quad & \{(n-1)S_1[n^2 - 3n + 3 - (n-1)b_2] \\ & -\frac{1}{4}(n-1)S_2[n^2 + 3n - 6 - (n^2 - n + 2)b_2] \\ & +S_0^2[n^2 - 3 - (n-1)^2 b_2]\}/n(n-2)(n-3)S_0^2 \end{aligned} \tag{7}$$

where S_0, S_1, and S_2 are as in the expressions above.

The interpretation of Geary's c is somewhat counterintuitive since it is based on a notion of dissimilarity (squared difference) rather than similarity. For values of c less than one, positive spatial autocorrelation is indicated, whereas negative spatial autocorrelation is suggested by values larger than one. As before, only when these values are significant can the null hypothesis of spatial randomness be rejected.

3.2.1 Local Geary

Similar to Moran's I, a local counterpart to Geary's c can be derived, using the general LISA principles outlined in Anselin (1995). The formal expression is:

$$c_i = \sum_{j \in J_i} w_{ij}(x_i - x_j)^2 \tag{8}$$

where, as before, J_i identifies the set of neighbors of observation i. Inference is based on conditional permutation (see also Sokal et al. 1998, for further technical details and applications).

3.3 Join Count Statistics

When the attribute variable is binary, that is,

$$x_i = \begin{cases} 1 & \text{if an event occurred in location } i \\ 0 & \text{if an event did not occur in location } i \end{cases}$$

a different methodological approach is needed to measure spatial autocorrelation, since both Moran's I and Geary's c require a ratio scale variable. A spatial autocorrelation statistic for binary data can be derived from the principles of Mantel's general cross product statistic, or Gamma statistic, $\Gamma = \sum_i \sum_j a_{ij} b_{ij}$, where a_{ij} and b_{ij} are matching elements in two matrices of similarity. This principle can be applied in the spatial case by considering one matrix as a matrix of attribute similarity and the other as a matrix of spatial similarity (i.e., the weights matrix W). The corresponding spatial statistic would then be $\Gamma = \sum_i \sum_j a_{ij} w_{ij}$.

In the case of a binary variable, three measures of attribute similarity (or dis-

similarity) can be used: $x_i x_j$, $(1-x_i)(1-x_j)$, and $(x_i - x_j)^2$. For $x_i x_j$, the product is non-zero when both x_i and x_j equal 1. Similarly, when both x_i and x_j are zero, the product $(1-x_i)(1-x_j)$ is non-zero. Both instances suggest positive spatial autocorrelation, i.e., similarity between the values at i and j. In contrast, when x_i and x_j take on different values (e.g., one is zero and the other one), then the expression $(x_i - x_j)^2$ is non-zero. This suggests negative spatial autocorrelation.

More specifically, this results in three so-called join count statistics for spatial autocorrelation among binary variables. Positive spatial autocorrelation is indicated by the BB (Black-Black, for a value of 1 in each location) and WW (White-White, for a value of 0 in each location) statistics. Formally:

$$BB = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} x_i x_j, \tag{9}$$

and,

$$WW = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (1-x_i)(1-x_j). \tag{10}$$

Negative spatial autocorrelation is indicated by the BW (Black-White, for a 1-0 pair) statistic. Formally:

$$BW = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2. \tag{11}$$

As shown in Cliff and Ord (1981), inference can be based on either an analytical or a computational approach. The moments of the statistics can be derived under a particular sampling model, with or without replacement. The permutation approach is more robust and is typically used in practice. For example, if there are n_1 observations with a value of 1 (B) in a sample of n, then the remaining

$n - n_1$ take on a value of 0 (W). The null hypothesis of no spatial autocorrelation can then be tested by permuting the n_1 black and the $n - n_1$ white cells over the observations.

3.3.1 Local Spatial Autocorrelation in Binary Data

The principle of a local spatial autocorrelation statistic can be extended to join count tests. In general, any statistic of the cross product form $\Gamma = \sum_i \sum_j a_{ij} w_{ij}$ can be localized for observation i as $\Gamma_i = \sum_j a_{ij} w_{ij}$ (Anselin 1995). In the case of a BB statistic, assuming i takes on the value of 1, this would take on the form:

$$BB_i = x_i \sum_j w_{ij} x_j. \tag{12}$$

Inference could be based on conditional permutation, although an analytical solution is straightforward. Under the null hypothesis of spatial randomness, each observation has equal probability of taking on a value of 1 (B). If p_b is the proportion of B values in the study region of size n, then the probability of finding x black observations in a subregion of n_i cells is simply given by the binomial distribution:

$$\Pr(X = x) = \binom{n_i}{x} p_b^x (1 - p_b)^{n_i - x}. \tag{13}$$

To be precise, the probability of "remaining" black observations should be adjusted for the fact that the location i takes on a black value, such that $p_b = n_b/(n-1)$. In practice, this adjustment will be negligible in any but the smallest data sets.

Given the total number of neighbors for an observation, it becomes a straightforward computation to determine how many black neighbors are needed for the

BB_i statistic to be significant. For example, locations for which $\Pr(X \geq x)$ is less than a chosen significance level (say p) would suggest the presence of a cluster.

However, unlike the case for ratio variables, where the average of the neighboring values is the only meaningful information that can be contained in the local statistic, the situation is more complex for discrete variables. The number of black neighboring cells can be arranged in several different ways (unless all neighbors take on the same value), some of which are more akin to the notion of a "cluster" than others. In this case, attention focuses not only on the first order neighbor relations between a location and its immediate neighbors, but also on the arrangement of the values between the neighbors themselves, which involves second order neighbor properties. In the literature, this distinction is referred to as *composition* and *configuration* (Li and Reynolds 1993, 1994, 1995).

This issue is further explored in a recent paper by Boots (2003). This is set in the context of 0-1 variables observed on a regular grid, which facilitates some of the formal derivations. Boots (2003) notes that when considered locally, configuration should be analyzed conditional on composition, i.e., after the number of a particular type of cell (black or white) has been accounted for. In his approach, local composition is measured by counting the number of cells of a particular type, while local configuration is measured by join counts. This requires a consideration of more than first order neighbors. In fact, he considers square windows of increasing size centered on the observation of interest (he uses square windows of dimension 3×3, 5×5 and 7×7). In other words, his notion of local statistic differs from the local join count expression given in Equation (12), but is more akin to the notion of a regional statistic, computed from a moving window. Boots (2003) outlines a classification into eight classes based on combinations of com-

position and configuration.

While quantifying composition is straightforward, obtaining a formal measure of configuration is more complex. One could express the total number of join counts in a sub-region as a proportion of all joins in that sub-region. Then a one-sample difference of proportion test could be employed to test the difference between this (local) proportion and the corresponding proportion over the entire study area. However, this approach is only valid when the composition of the sub-region and that of the entire study area are the same. The proper way to test for local configuration should thus be conditional on composition. In other words, given the number of black cells in a sub-region, the number of joins needs to be assessed relative to a random assignment. Analytical approaches towards inference are limited by the small number of cells in the moving window. Boots (2003) enumerates all possible cases for the three window sizes considered, up to 7×7. However, this approach is limited to a regular grid configuration.

To extend these ideas to an irregular spatial configuration would likely require a permutation approach. In addition, the notion of a moving window would need to be formalized. Obvious approaches would be to use either higher orders of contiguity, distance bands, or nearest neighbor criteria. This remains to be further explored and is not applied in our empirical study.

4 Data and Software

Before moving on to the empirical study of violent events in Baghdad, Iraq, we briefly review the characteristics of the data and the software used in the analysis.

4.1 Data

Figure 6 shows the city of Baghdad by neighborhood. Due to data limitations, the study only considers 45 neighborhood areas in the city of Baghdad. Figure 7 presents the point locations of these 45 neighborhoods (marked by their sequence number). These numbers correspond to the list of neighborhood names in Appendix A. The use of points rather than polygons to represent the neighborhoods avoids problems with the precise delineation of the boundaries between them and provides sufficient locational detail to allow for the computation of both global and local spatial autocorrelation statistics. In order to visualize the point data results in map format for this report, the 45 points are converted to (Thiessen) polygons to create a neighborhood around each point, whereby neighborhood boundaries are defined by the median distance between closest points. Figure 7 presents these 45 neighborhoods as an overlay of the original Baghdad neighborhoods in Figure 6.[1]

[1] To distinguish between the neighborhoods of both figures, the maps frequently refer to the 45 neighborhoods of this report as ”incident areas.”

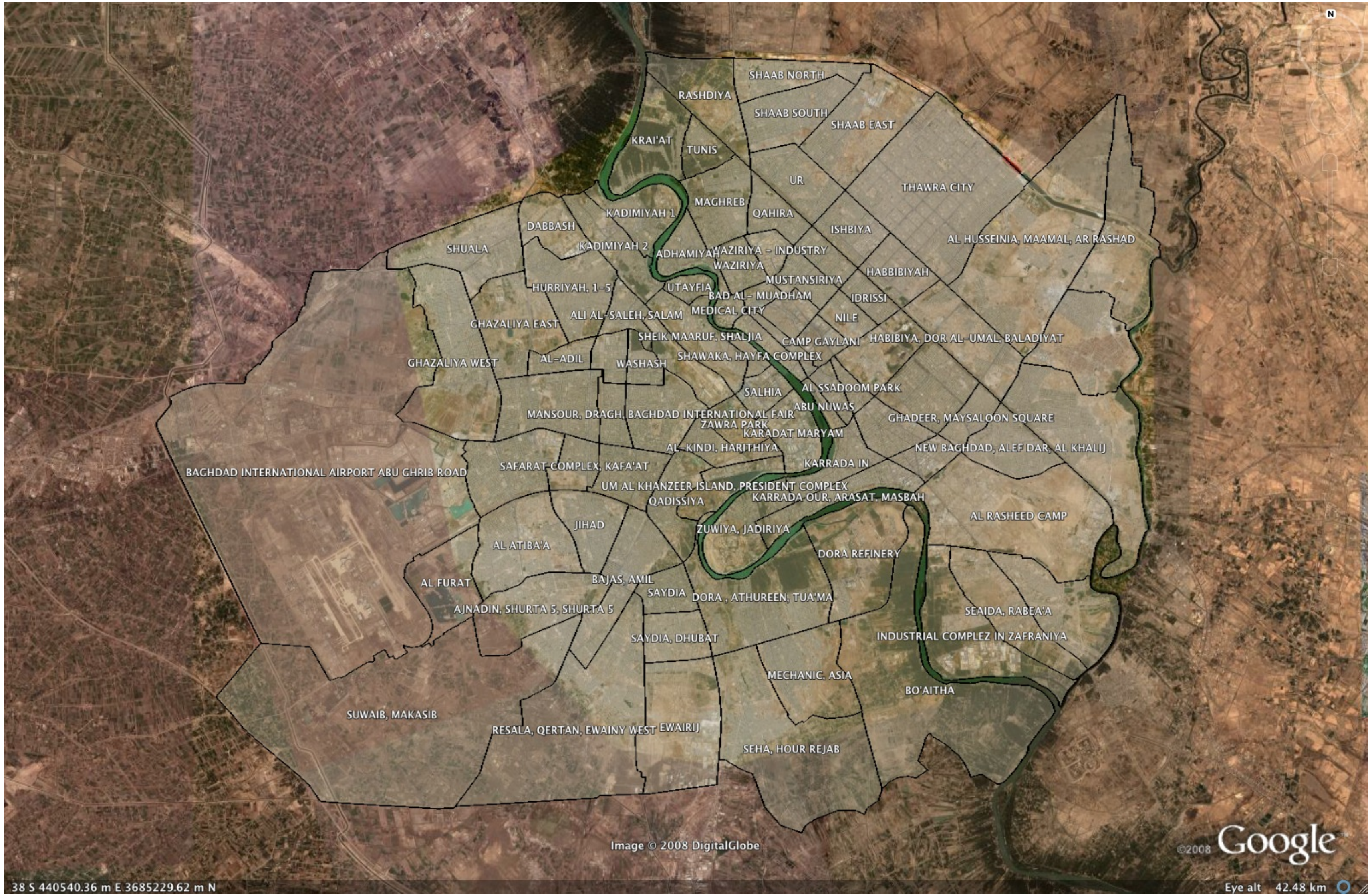


Figure 6: Baghdad Neighborhoods

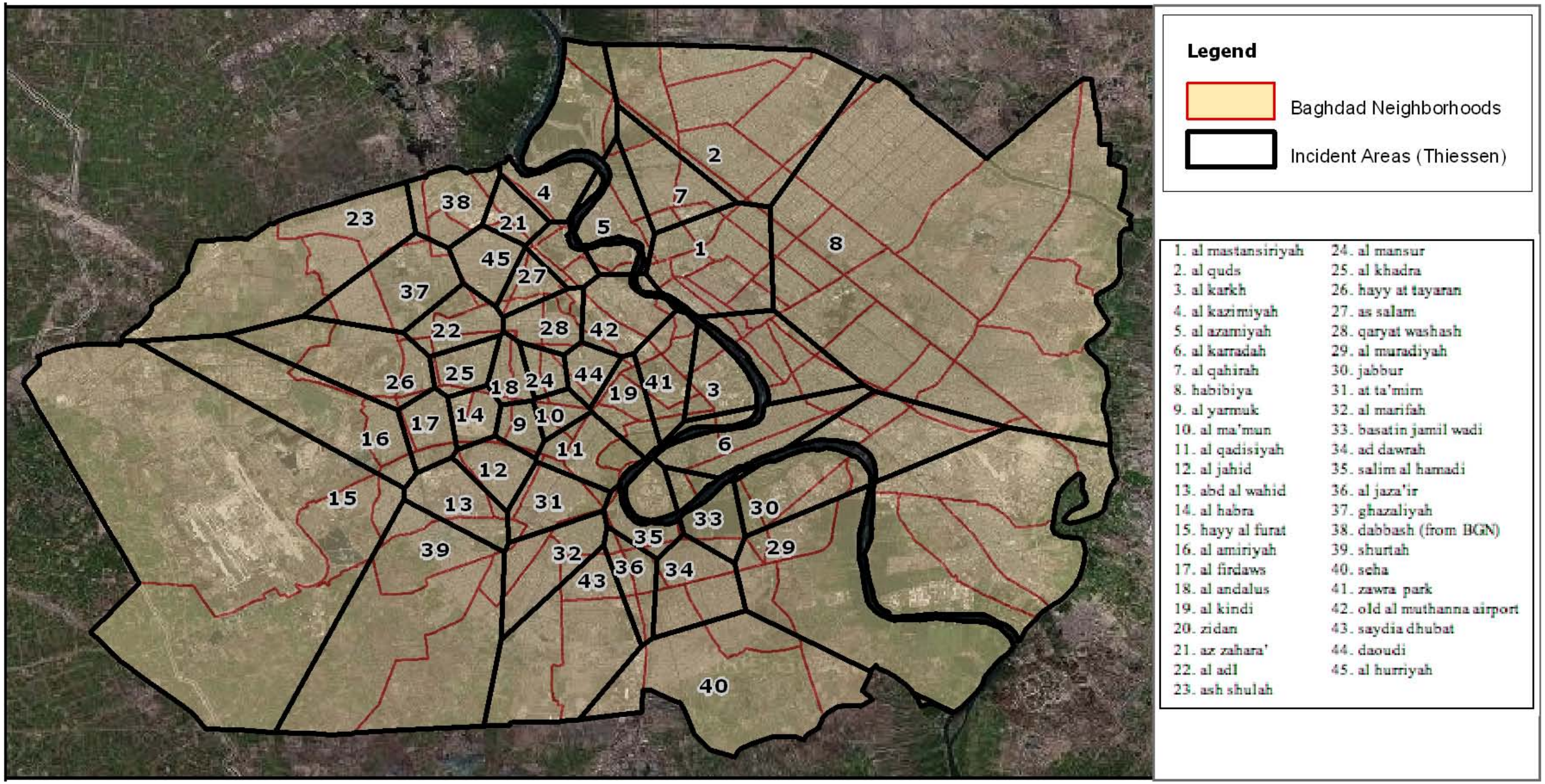


Figure 7: 45 Point Locations and Surrounding Neighborhoods

The original source for the data is a content analysis of news reports available as open source. This involved a coding system that classified events reported on in the news sources. While precise geocoding was not possible, there was sufficient information in the reports to associate each event with a neighborhood.

A detailed coding system was used to classify the events. This system has a nested structure, going from more general descriptions to more precise categories. The structure is such that all events reported in a lower level (more specific) category are also included in the total at the higher level. However, for some events, it was not possible to find a more specific classification, so that the higher level total does not necessarily correspond to the sum of all lower level categories. For example, if the variable 7620IZ_suicide_bombing is non-zero for one of the 45 neighborhoods, the higher level variable 762_Suicide should also be non-zero. However, the reverse is not necessarily the case, i.e., the count of events at the higher level could be larger than (or equal to) the sum of events at lower level categories. The detailed classification is given in Appendix B.

In the spatial analysis that follows in Section 5, we will report results for a spatial weights matrix based on four nearest neighbors.[2] The structure of dependence between observations implied by this weights structure is illustrated in Figure 8. Each point represents a spatial unit, with the red lines connecting each observation to its four nearest neighbors. The denseness of the resulting network highlights the number of connections between the sample points.

The content analysis of news sources yielded 112 variables for which at least one event was recorded. However, a closer analysis of the frequency distribution of the number of neighborhoods for which an event was observed for each of the

[2]Other spatial weights were considered as well without substantively affecting the results.

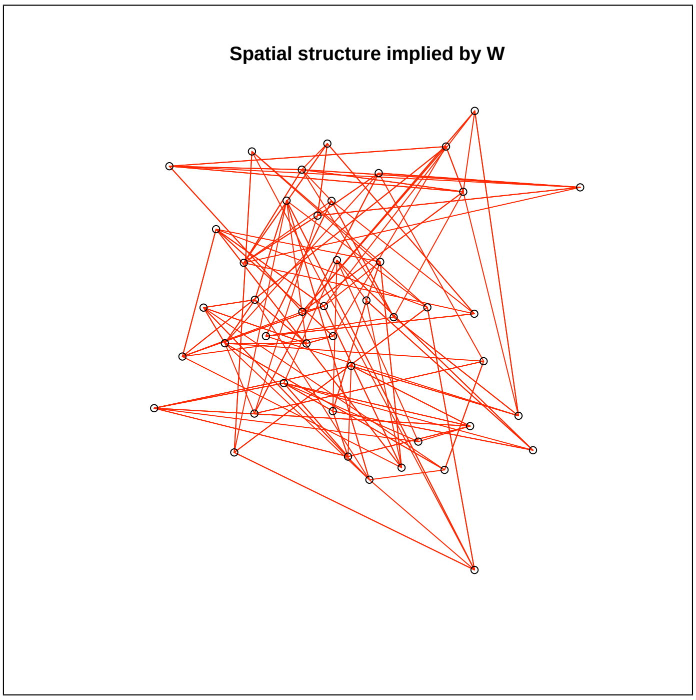


Figure 8: Structure of Dependence Implied by a 4-Nearest Neighbors Weights Matrix.

variables reveals a highly skewed result. As shown in Table 1, 40 variables only record events in a single neighborhood, and 19 variables are only present in two neighborhoods. Across all variables, the average number of neighborhoods with recorded events amounts to four. For more than 75% of the variables, there are fewer than six neighborhoods with recorded events.

This result could be due to the detailed nature of the classification which yields insufficient (rare) events over the course of a year. Alternatively, it could also point to a problem with the spatial scale of analysis. With relatively large (in area) neighborhoods as units of analysis, it is conceivable that highly spatially concentrated events are only recorded for one or two neighborhoods. Such events may show spatial pattern within the neighborhood scale, but our analysis would not be able to detect such pattern.

In order for meaningful spatial analysis to be carried out, we limited our scope to those variables for which at least four different neighborhoods (out of 45) recorded events. This eliminated 80 variables from consideration, leaving 32 to be included in the analysis.

4.2 Software

The analysis that follows was largely performed using the R statistical software. R is free software that can be downloaded from the CRAN website (at the address http://cran.r-project.org/). R is a collaborative project with many contributors. Most of the analysis presented here will be based on functions that are supplied as *packages*. The **spdep** library (Bivand 2001, 2006, 2002, Bivand and Gebhardt 2000, Bivand and Portnov 2004) for spatial regression analysis contains most of

Table 1: Distribution of Non-Zero Observations for Each Variable.

Non-zero	Number
1	40
2	19
3	14
4	7
5	6
6	4
7	4
8	4
9	3
10	4
11	1
15	5
21	1
TOT	112

the tools needed to perform statistical testing for the presence of spatial autocorrelation.

In addition, some of the methods discussed in Section 2 required additional computer programming, since they are not currently available in any software package (e.g., the local measures for binary data). When needed, these computer programs were also written in the R language.

Finally, the spatial weights matrix was constructed from a shape file using GeoDa software. GeoDa is a free downloadable software package (http://geodacenter.asu.edu/software/downloads) developed at the GeoDa center for Geospatial Analysis and Computation at the Arizona State University (Anselin et al. 2006). GeoDa was also used to implement the LISA analysis.

5 Spatial Analysis

In this section, we present results for the analysis of patterns in the distribution of violent events in Baghdad using statistics for global and local spatial autocorrelation. We proceed in two different ways. First, we consider the data as counts of events, but ignore the specific count nature of the data (e.g., as would be the case using a Poisson model). In other words, we treat the counts as if they were ratio scale variables. While a loss of precision will be associated with this approach, it does provide a qualitative measure of the degree of spatial clustering in the data. In a second approach, we convert the counts to a binary 0-1 scale and apply join count statistics to the resulting data. In both instances, we take a global as well as a local perspective, applying the statistical tests outlined in detail in Section 3.

5.1 Counts of Events

5.1.1 Global Spatial Autocorrelation

Moran's I

We begin the analysis of global spatial autocorrelation by considering the Moran's I statistic using four nearest neighbor spatial weights. The main results are reported in Tables 2 and 3, respectively for the normal and randomization assumption using an analytical approach towards inference.

Each table contains four columns. The first column reports the value of the sample estimate of Moran's I, the second column is the expected value, then the variance and, finally, an indication of the significance level for a two-sided statistical test. The difference between the two tables is only in the fourth and last columns, since only the variance under the null differs between the two ap-

proaches. The statistic itself and its expected value are the same.

Of the 32 variables considered in Table 2, only 166_1660IZ (within city inmigration, hayy to hayy) is highly significant ($p < 0.0001$) and positive. This suggests a significant degree of clustering at the neighborhood level. Note that from the global statistic, it is not possible to assess whether this is clustering of high values or low values, only of "similar" values. None of the other variables showed a significant level of spatial autocorrelation, not even at a p levels of 0.05 or 0.10.

The results in Table 3 are largely the same. Again, the test statistics for variable 166_1660IZ is found to be highly significant ($p < 0.0001$) and positive. In addition, but only at a very conservative significance level of $p < 0.10$, two new variables show weak evidence of positive spatial autocorrelation: 627_6277IZ_6820IZ (bombing by other means – killed) and 627_6277IZ_6821IZ (bombing by other means – wounded). This suggests some clustering of similar values of these variables at the neighborhood scale.

A final assessment of Moran's I is carried out by means of a permutation test. In Figure 9, a graph with the reference distribution constructed from randomly permuted data sets is shown for each variable. The reference distribution was based on 999 independent samples. The density function (black curve in the graphs) corresponds to the distribution of the statistic under the null. The red vertical line is drawn at the observed value of Moran's I for that variable. A pseudo significance test can be based on how extreme the observed value is relative to the reference distribution. For variable 166_1660IZ (within city inmigration, hayy to hayy), there is again strong evidence of positive spatial autocorrelation. Similar to the results under randomization, there is weak evidence of positive spatial au-

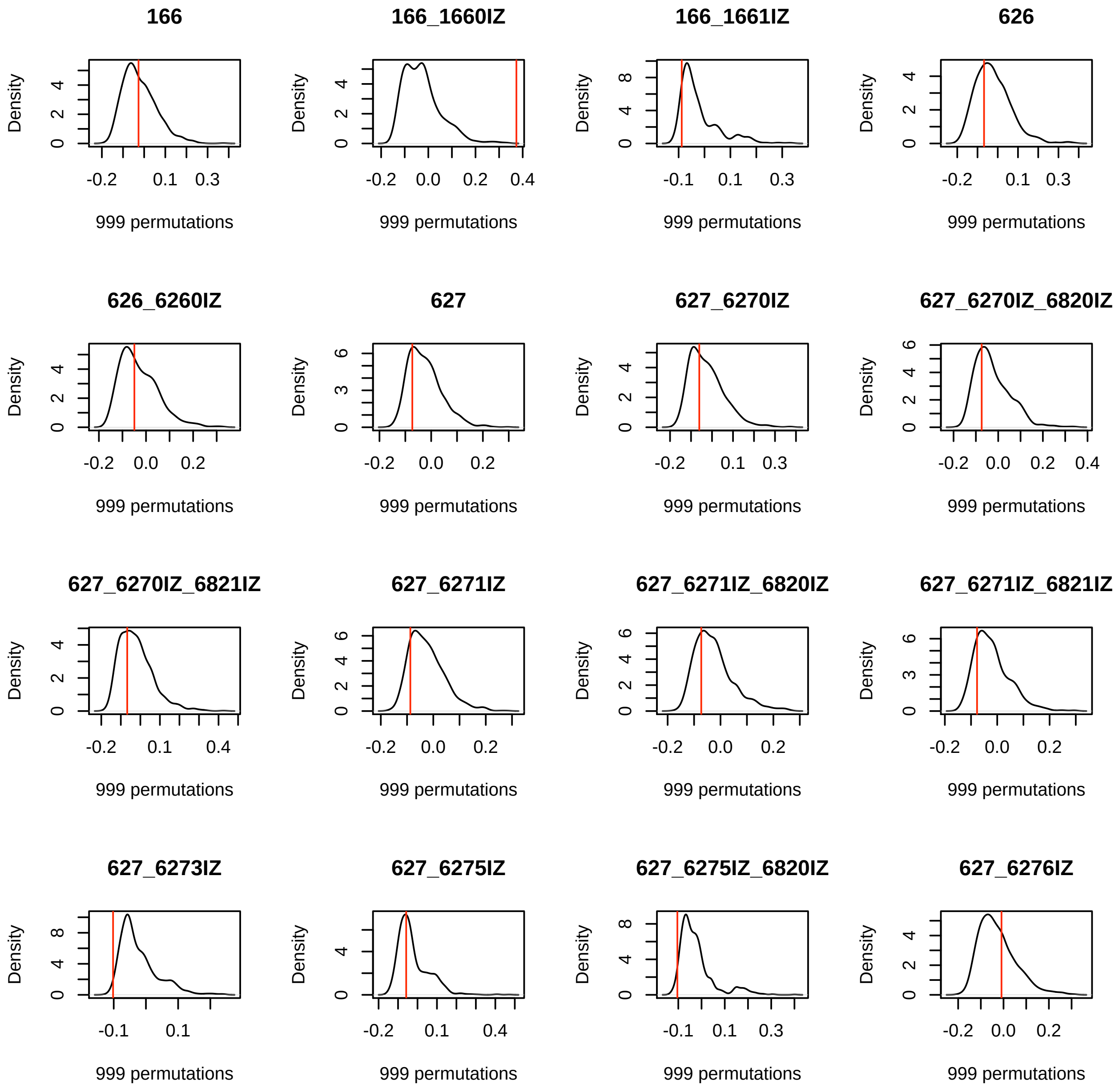
tocorrelation for the additional variables 627_6277IZ_6820IZ (bombing by other means – killed) and 627_6277IZ_6821IZ (bombing by other means – wounded).

Table 2: Moran's I under Normality.

	I	E(I)	Var(I)	p-value
166_1660IZ (within city inmigration, hayy to hayy)	0.373	-0.023	0.009	0.000
627_6277IZ_6820IZ (bombing by other means - killed)	0.115	-0.023	0.009	0.137
627_6277IZ_6821IZ (bombing by other means - wounded)	0.106	-0.023	0.009	0.163
627_6275IZ_6820IZ (attack on political officials - killed)	-0.104	-0.023	0.009	0.379
627_6273IZ (firing on crowd)	-0.102	-0.023	0.009	0.391
627_6277IZ (bombing by other means)	0.053	-0.023	0.009	0.413
682_6821IZ (offenses against life - wounded)	0.046	-0.023	0.009	0.455
727 (aftermath of combat)	-0.091	-0.023	0.009	0.458
727_7270IZ_5630IZ (collateral civilian wounded -sunni)	-0.091	-0.023	0.009	0.463
727_7270IZ (killed while captive)	-0.089	-0.023	0.009	0.475
166_1661IZ (within city outmigration, hayy to hayy)	-0.088	-0.023	0.009	0.480
627_6271IZ (roadside or on road bombing)	-0.087	-0.023	0.009	0.483
682_6820IZ_5631IZ (offenses against life - killed - sunni)	-0.087	-0.023	0.009	0.483
683 (offenses against the person)	-0.079	-0.023	0.009	0.545
683_6830IZ (kidnapping indigenous person)	-0.079	-0.023	0.009	0.545
627_6271IZ_6821IZ (roadside or on road bombing - wounded)	-0.077	-0.023	0.009	0.557
627_6270IZ_6820IZ (car bombing - killed)	-0.074	-0.023	0.009	0.579
627 (informal in group justice)	-0.073	-0.023	0.009	0.587
627_6271IZ_6820IZ (roadside or on road bombing - killed)	-0.073	-0.023	0.009	0.588
626 (social control)	-0.068	-0.023	0.009	0.626
627_6270IZ_6821IZ (car bombing - wounded)	-0.067	-0.023	0.009	0.630
627_6278IZ (rocket and or mortar attacks)	-0.061	-0.023	0.009	0.681
627_6270IZ (car bombing)	-0.061	-0.023	0.009	0.682
627_6276IZ_6820IZ (attack on mosque - killed)	-0.060	-0.023	0.009	0.683
627_6275IZ (attack on political officials)	-0.058	-0.023	0.009	0.702
626_6260IZ (written threat)	-0.049	-0.023	0.009	0.773
627_6278IZ_6821IZ (rocket and or mortar attacks - wounded)	-0.048	-0.023	0.009	0.784
627_6276IZ (attack on mosque)	-0.009	-0.023	0.009	0.881
627_6278IZ_6820IZ (rocket and or mortar attacks - killed)	-0.029	-0.023	0.009	0.949
682_6820IZ (offenses against life - killed)	-0.027	-0.023	0.009	0.964
166 (within city migration, hayy to hayy)	-0.027	-0.023	0.009	0.967
682 (offenses against life)	-0.020	-0.023	0.009	0.973

Table 3: Moran's I under Randomization.

	I	E(I)	Var(I)	p-value
166_1660IZ (within city inmigration, hayy to hayy)	0.373	-0.023	0.007	0.000
627_6277IZ_6821IZ (bombing by other means - wounded)	0.106	-0.023	0.004	0.055
627_6277IZ_6820IZ (bombing by other means - killed)	0.115	-0.023	0.006	0.067
627_6273IZ (firing on crowd)	-0.102	-0.023	0.004	0.178
627_6277IZ (bombing by other means)	0.053	-0.023	0.004	0.240
727 (aftermath of combat)	-0.091	-0.023	0.004	0.263
627_6275IZ_6820IZ (attack on political officials - killed)	-0.104	-0.023	0.005	0.274
727_7270IZ (killed while captive)	-0.089	-0.023	0.004	0.290
727_7270IZ_5630IZ (collateral civilian wounded -sunni)	-0.091	-0.023	0.005	0.330
627_6271IZ (roadside or on road bombing)	-0.087	-0.023	0.005	0.359
166_1661IZ (within city outmigration, hayy to hayy)	-0.088	-0.023	0.005	0.376
682_6821IZ (offenses against life - wounded)	0.046	-0.023	0.007	0.416
627_6271IZ_6821IZ (roadside or on road bombing - wounded)	-0.077	-0.023	0.005	0.431
627 (informal in group justice)	-0.073	-0.023	0.004	0.435
683 (offenses against the person)	-0.079	-0.023	0.005	0.450
683_6830IZ (kidnapping indigenous person)	-0.079	-0.023	0.005	0.450
682_6820IZ_5631IZ (offenses against life - killed - sunni)	-0.087	-0.023	0.008	0.459
627_6271IZ_6820IZ(roadside or on road bombing - killed)	-0.073	-0.023	0.005	0.488
627_6270IZ_6820IZ (car bombing - killed)	-0.074	-0.023	0.007	0.532
627_6278IZ (rocket and or mortar attacks)	-0.061	-0.023	0.004	0.535
627_6270IZ_6821IZ (car bombing - wounded)	-0.067	-0.023	0.007	0.591
626 (social control)	-0.068	-0.023	0.008	0.606
627_6270IZ (car bombing)	-0.061	-0.023	0.007	0.649
627_6276IZ_6820IZ (attack on mosque - killed)	-0.060	-0.023	0.007	0.653
627_6278IZ_6821IZ (rocket and or mortar attacks - wounded)	-0.048	-0.023	0.003	0.653
627_6275IZ (attack on political officials)	-0.058	-0.023	0.007	0.663
626_6260IZ (written threat)	-0.049	-0.023	0.007	0.754
627_6276IZ (attack on mosque)	-0.009	-0.023	0.007	0.870
627_6278IZ_6820IZ (rocket and or mortar attacks - killed)	-0.029	-0.023	0.004	0.922
682_6820IZ (offenses against life - killed)	-0.027	-0.023	0.003	0.941
682 (offenses against life)	-0.020	-0.023	0.004	0.959
166 (within city migration, hayy to hayy)	-0.027	-0.023	0.007	0.962


166
166_1660IZ
166_1661IZ
626
626_6260IZ
627
627_6270IZ
627_6270IZ_6820IZ
627_6270IZ_6821IZ
627_6271IZ
627_6271IZ_6820IZ
627_6271IZ_6821IZ
627_6273IZ
627_6275IZ
627_6275IZ_6820IZ
627_6276IZ
Density
999 permutations

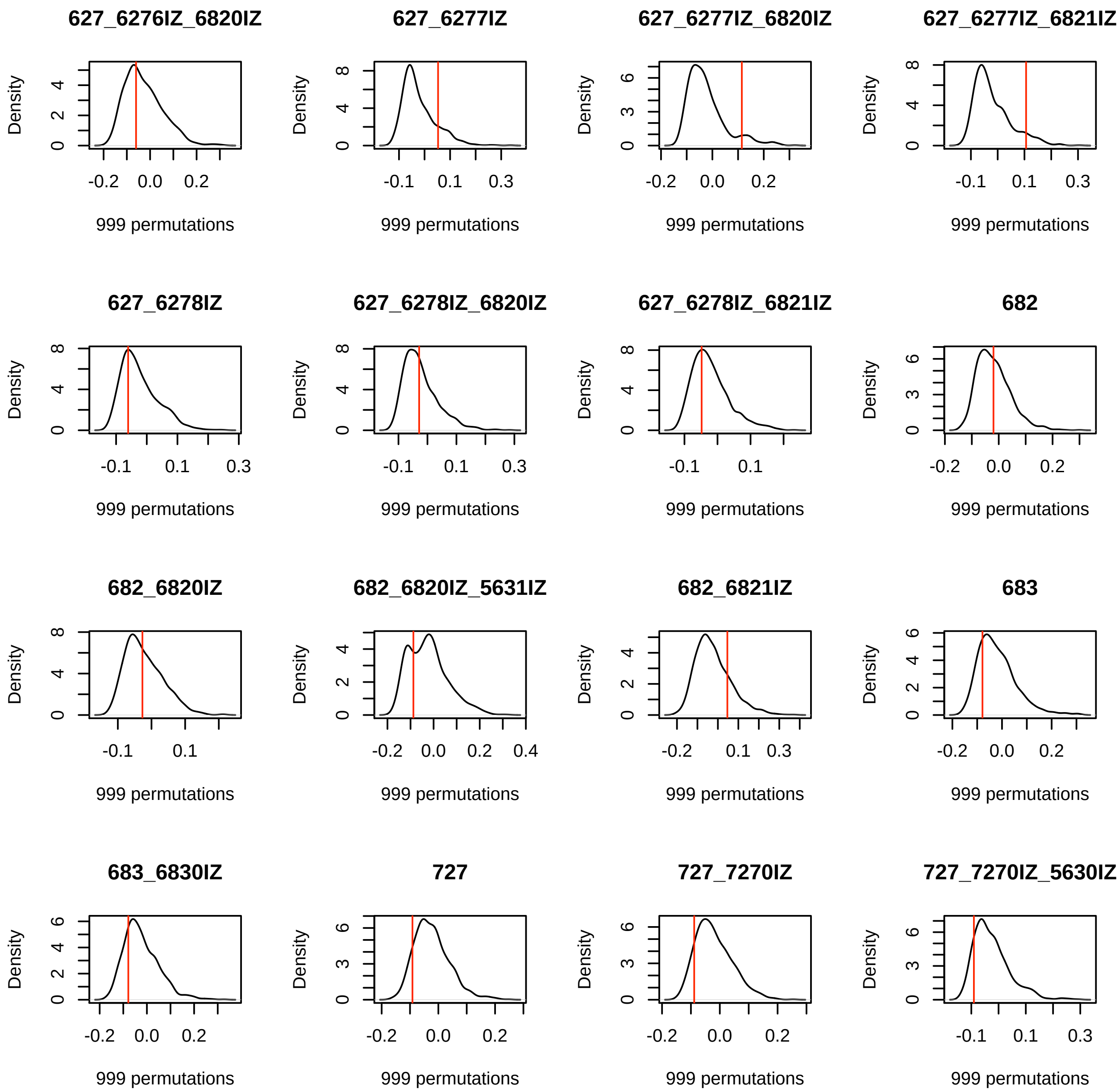


Figure 9: Moran's I under Permutation Approach, Number of Permutations is 999.

Geary's c

Table 4 summarizes the results from the calculation of the Geary's c statistics for spatial autocorrelation. The table consists of four columns. The first column contains the value of Geary's c statistic. The second column reports the variance calculated under the assumption of randomization (Equation 7).[3] The expected value is equal to one. The standard deviate is calculated by subtracting the expected value from the value of the statistics and dividing by the standard deviation. The last column contains an indication of the significance of a two-sided statistical test. Note that the interpretation of Geary's c is different from that for Moran's I. Positive spatial autocorrelation is obtained for $c < 1$, negative spatial autocorrelation for $c > 1$.

The results provide quite a different portrayal of the spatial structure in the data. In terms of positive spatial autocorrelation, the same variable as before (166_1660IZ – within city inmigration, hayy to hayy) is highly significant ($p = 0.0008$). However, this is the only variable with a Geary c statistic less than one. Of the remaining variables, 21 show significant negative spatial autocorrelation at $p < 0.10$, of which seven are significant at $p < 0.05$.[4]

The seven variables with the most significant negative spatial autocorrelation are 627_6273IZ (firing on crowd, $p < 0.03$), 627_6270IZ_6820IZ (car bombing, killed, $p < 0.04$), 627_6270IZ_6821IZ (car bombing, wounded, $p < 0.04$), 627_6271IZ_6821IZ (road side or road bombing, wounded, $p < 0.04$), 683 (offenses against the person, $p < 0.04$), 683_6830IZ (kidnapping indigenous person, $p < 0.04$), and 727_7270IZ_5630 (killed while captive, Shia, $p < 0.05$).

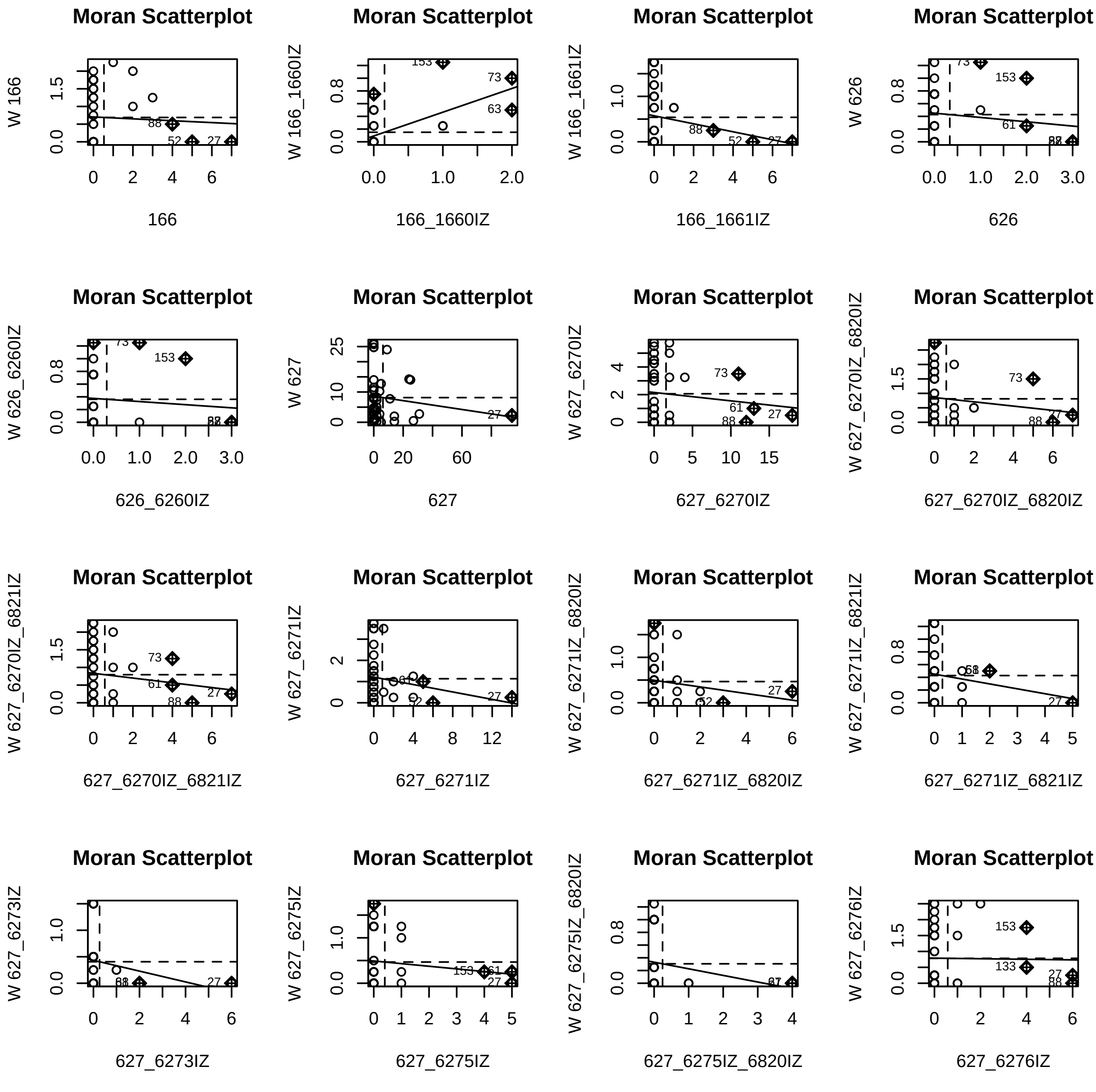
[3] Inference based on normality produced identical results and therefore it is not reported.

[4] In general, because it is based on the squared difference, a measure of dissimilarity, Geary's c tends to have more power to detect negative spatial autocorrelation.

Table 4: Geary's c under Randomization.

	c	Var(c)	Standard deviate	p-value
166_1660IZ (within city inmigration, hayy to hayy)	0.5549	0.0175	-3.3626	0.0008
627_6273IZ (firing on crowd)	1.3964	0.0312	2.2447	0.0248
627_6270IZ_6821IZ (car bombing - wounded)	1.2830	0.0171	2.1622	0.0306
627_6270IZ_6820IZ (car bombing - killed)	1.2863	0.0177	2.1531	0.0313
627_6271IZ_6821IZ (roadside or on road bombing - wounded)	1.3419	0.0258	2.1291	0.0332
683 (offenses against the person)	1.3207	0.0227	2.1263	0.0335
683_6830IZ (kidnapping indigenous person)	1.3207	0.0227	2.1263	0.0335
727_7270IZ_5630IZ (collateral civilian wounded -sunni)	1.3245	0.0255	2.0340	0.0420
627_6270IZ (car bombing)	1.2520	0.0168	1.9457	0.0517
627 (informal in group justice)	1.3251	0.0284	1.9305	0.0535
627_6271IZ (roadside or on road bombing)	1.2987	0.0249	1.8925	0.0584
627_6275IZ_6820IZ (attack on political officials - killed)	1.2823	0.0226	1.8762	0.0606
727 (aftermath of combat)	1.3212	0.0300	1.8559	0.0635
627_6276IZ_6820IZ (attack on mosque - killed)	1.2360	0.0165	1.8358	0.0664
727_7270IZ (killed while captive)	1.3137	0.0294	1.8284	0.0675
166_1661IZ (within city outmigration, hayy to hayy)	1.2770	0.0231	1.8235	0.0682
627_6276IZ (attack on mosque)	1.2288	0.0160	1.8062	0.0709
627_6278IZ (rocket and or mortar attacks)	1.2984	0.0300	1.7240	0.0847
627_6278IZ_6821IZ (rocket and or mortar attacks - wounded)	1.3017	0.0324	1.6756	0.0938
627_6271IZ_6820IZ (roadside or on road bombing - killed)	1.2566	0.0239	1.6587	0.0972
682_6820IZ (offenses against life - killed)	1.2973	0.0324	1.6505	0.0988
626 (social control)	1.1955	0.0141	1.6466	0.0996
682 (offenses against life)	1.2692	0.0295	1.5679	0.1169
166 (within city migration, hayy to hayy)	1.1965	0.0178	1.4737	0.1406
626_6260IZ (written threat)	1.1828	0.0155	1.4669	0.1424
627_6278IZ_6820IZ (rocket and or mortar attacks - killed)	1.2510	0.0305	1.4384	0.1503
627_6277IZ (bombing by other means)	1.2011	0.0283	1.1952	0.2320
627_6275IZ (attack on political officials)	1.1483	0.0182	1.0976	0.2724
682_6820IZ_5631IZ (offenses against life - killed - sunni)	1.1275	0.0139	1.0826	0.2790
627_6277IZ_6821IZ (bombing by other means - wounded)	1.1242	0.0269	0.7577	0.4487
627_6277IZ_6820IZ (bombing by other means - killed)	1.0916	0.0223	0.6140	0.5392
682_6821IZ (offenses against life - wounded)	1.0442	0.0157	0.3534	0.7238

Given the small number of non-zero observations, this would suggest that neighborhoods with these types of events tend to be surrounded by neighborhoods without events, although this needs to be further investigated by means of local statistics.


Moran Scatterplot
W 166
166
Moran Scatterplot
W 166_1660IZ
166_1660IZ
Moran Scatterplot
W 166_1661IZ
166_1661IZ
Moran Scatterplot
W 626
626
Moran Scatterplot
W 626_6260IZ
626_6260IZ
Moran Scatterplot
W 627
627
Moran Scatterplot
W 627_6270IZ
627_6270IZ
Moran Scatterplot
W 627_6270IZ_6820IZ
627_6270IZ_6820IZ
Moran Scatterplot
W 627_6270IZ_6821IZ
627_6270IZ_6821IZ
Moran Scatterplot
W 627_6271IZ
627_6271IZ
Moran Scatterplot
W 627_6271IZ_6820IZ
627_6271IZ_6820IZ
Moran Scatterplot
W 627_6271IZ_6821IZ
627_6271IZ_6821IZ
Moran Scatterplot
W 627_6273IZ
627_6273IZ
Moran Scatterplot
W 627_6275IZ
627_6275IZ
Moran Scatterplot
W 627_6275IZ_6820IZ
627_6275IZ_6820IZ
Moran Scatterplot
W 627_6276IZ
627_6276IZ

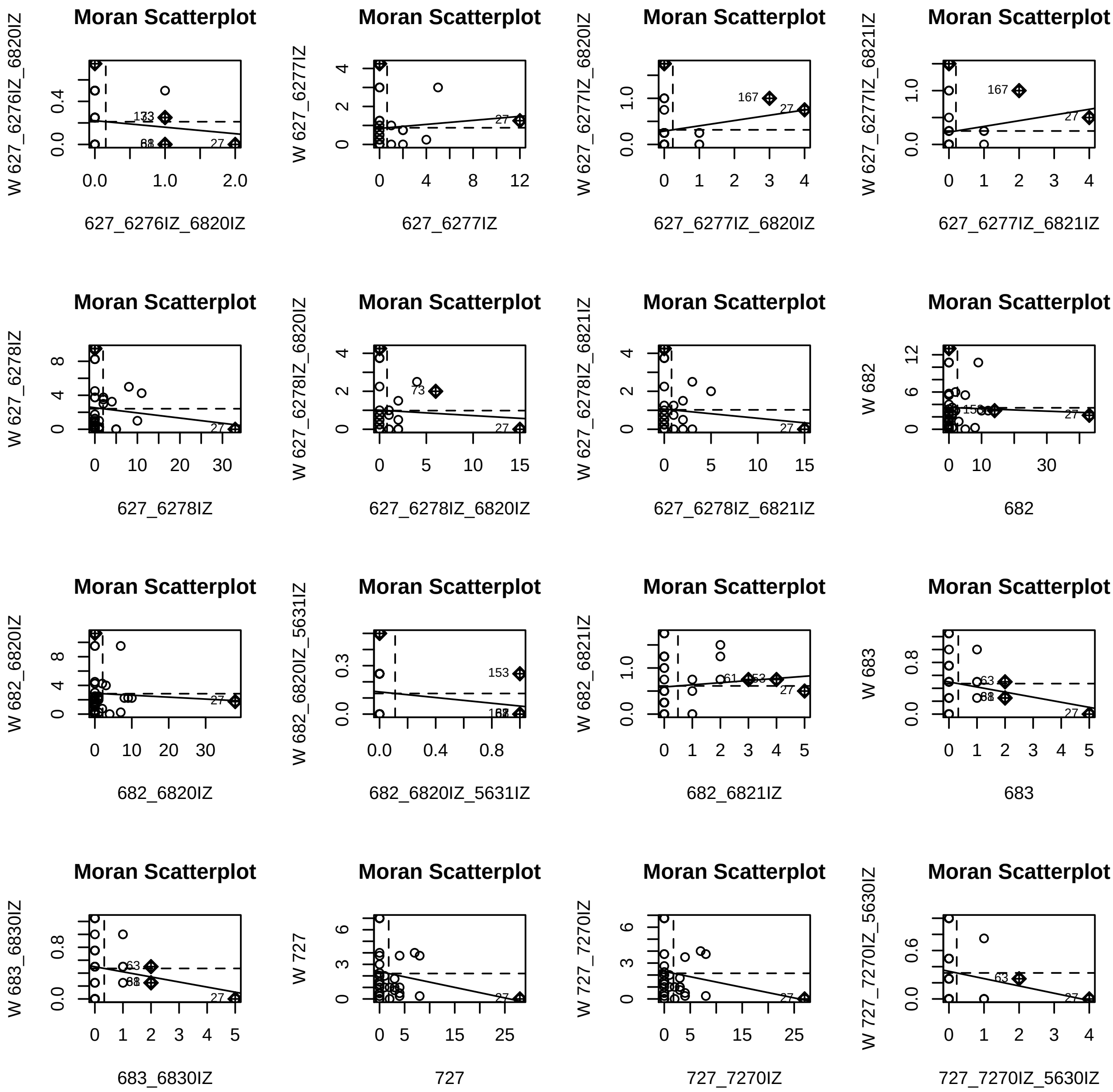


Figure 10: Moran Scatterplot

Moran Scatterplot

A final investigation of the extent of global spatial autocorrelation is carried out by means of the Moran Scatterplot. In Figure 10, this is illustrated for all the variables considered (again, using 4 nearest neighbor spatial weights).

The graphs illustrate the problems caused by the paucity of observations. In most of them, there is a high concentration of zero values, suggested by the vertical linear pattern in the left of the graph. Consequently, the slope of the linear smoother is highly influenced by a few (outlying) observations and is not a very reliable indicator of overall pattern. For the same variables identified before, we find a positive slope. For 166_1660IZ (within city inmigration, hayy to hayy) this is most pronounced. For the other two variables, 627_6277IZ_6820IZ (bombing by other means – killed) and 627_6277IZ_6821IZ (bombing by other means – wounded) the evidence is much weaker. In most of the other cases, the linear regression is fairly flat. In several instances, it is negative, although often driven by very few observations.

The slope as such does not provide any indication of significance, but it provides a suggestion of the overall trend. The graph also allows a cursory investigation of the type of spatial association (high-high, low-low, high-low, and low-high), which will be further investigated by means of local statistics in Section 5.1.2.

In sum, the findings from the global spatial autocorrelation analysis suggest that there is little overall evidence of spatial clustering. Of the 32 variables considered, only one consistently shows significant and positive spatial autocorrelation. Only two others show weak positive spatial autocorrelation. Several others show negative spatial autocorrelation, but only for Geary's c statistic. As mentioned

before, in part this may be due to the spatial scale of analysis. The spatial autocorrelation tests considered take into account the similarity between neighboring spatial units. If the interesting patterns occur within the neighborhood unit (and not between them), they will not be detected at this scale of analysis.

5.1.2 Local Spatial Autocorrelation

We now turn to the identification of local clusters and spatial outliers by means of the Local Moran statistic. Table 5 summarizes the results. It should be kept in mind that for some variables, these analyses are based on a very small number of non-zero observations, which may lead to unreliable results. Also, the Local Moran is applied without consideration of the count nature of the data, nor of the presence of zero values. Inference is based on the conditional permutation method using 999 replications.

For all but one of the variables at least one location was identified with a significant local Moran at $p < 0.01$. The number of locations ranged from one (for four variables) to five (for four other variables). Striking, but not surprising, is the general evidence of negative local spatial autocorrelation or spatial outliers (predominantly high-low) (Figure 11 and Table 6). As pointed out in the discussion of Geary's c, the sparseness of violent events will tend to result in a spatial distribution where a neighborhood with one or more events will be surrounded by neighborhoods without events. This is likely to lead to a significant negative local Moran statistic. Only for 166_1660IZ (within city inmigration, hayy to hayy) do we again find evidence of positive local spatial autocorrelation, although a spatial outlier is identified as well (Figure 12).

Three locations appear for 17 of the variables: #16 (al amiriyah) #34 (ad

Table 5: Local Moran by Type of Spatial Correlation. HH (High-High), HL (High-Low), LH (Low-High)

obs	I	Type	p-value
	166 (within city migration, hayy to hayy)		
al amiriyah (16)	-1.10	HL	0.001
ad dawrah (34)	-1.59	HL	0.001
	166_1660IZ (within city inmigration, hayy to hayy)		
ash shulah (23)	6.91	HH	0.001
ghazaliyah (37)	4.10	HH	0.001
dabbash (38)	-0.41	LH	0.008
	166_1661IZ (within city outmigration, hayy to hayy)		
al amiriyah (16)	-0.98	HL	0.001
ad dawrah (34)	-1.40	HL	0.001
	626 (social control)		
al amiriyah (16)	-1.22	HL	0.001
ad dawrah (34)	-1.22	HL	0.001
al hurriyah (45)	-1.22	HL	0.001
	626_6260IZ (written threat)		
al amiriyah (16)	-1.18	HL	0.001
al mansur (24)	-0.31	HL	0.001
ad dawrah (34)	-1.18	HL	0.001
al hurriyah (45)	-1.18	HL	0.001
	627_6270IZ (car bombing)		
al amiriyah (16)	-0.04	HL	0.001
shurtah (39)	-0.04	HL	0.001
al hurriyah (45)	-1.04	HL	0.001

	627_6270IZ_6820IZ (car bombing - killed)		
al amiriyah (16)	-0.10	HL	0.001
shurtah (39)	-0.10	HL	0.001
al hurriyah (45)	-1.33	HL	0.001
	627_6270IZ_6821IZ (car bombing - wounded)		
shurtah (39)	-0.11	HL	0.001
al hurriyah (45)	-1.14	HL	0.001
	627_6271IZ (roadside or on road bombing)		
al amiriyah (16)	-0.74	HL	0.001
	627_6271IZ_6820IZ (roadside or on road bombing - killed)		
al amiriyah (16)	-0.86	HL	0.001
as salam (27)	-0.20	HL	0.001
salim al hamadi (35)	-0.45	LH	0.008
ghazaliyah (37)	-0.53	HL	0.001
	627_6271IZ_6821IZ (roadside or on road bombing - wounded)		
al amiriyah (16)	-0.27	HL	0.001
ad dawrah (34)	-1.80	HL	0.001

	627_6273IZ (firing on crowd)		
al mansur (24)	-0.48	HL	0.001
ad dawrah (34)	-1.57	HL	0.001
al hurriyah (45)	-0.48	HL	0.001
	627_6275IZ (attack on political officials)		
al amiriyah (16)	-0.17	HL	0.001
ad dawrah (34)	-1.29	HL	0.001
	627_6275IZ_6820IZ (attack on political officials - killed)		
al qadisiyah (11)	-0.25	HL	0.001
al amiriyah (16)	-0.25	HL	0.001
al mansur (24)	-1.25	HL	0.001
ad dawrah (34)	-1.25	HL	0.001
ghazaliyah (37)	-0.25	HL	0.001
	627_6276IZ (attack on mosque)		
al amiriyah (16)	-0.11	HL	0.001
al mansur (24)	-0.11	HL	0.001
al hurriyah (45)	-1.42	HL	0.001
	627_6276IZ_6820IZ (attack on mosque - killed)		
al mansur (24)	-0.73	HL	0.001
ad dawrah (34)	-1.60	HL	0.001
al hurriyah (45)	-0.73	HL	0.001
	627_6277IZ (bombing by other means)		
al amiriyah (16)	-0.21	HL	0.001
salim al hamadi (35)	-0.57	LH	0.001
al jaza' ir (36)	-0.57	LH	0.002
seha (40)	-0.57	LH	0.001
al hurriyah (45)	-0.06	HL	0.001

	627_6277IZ_6820IZ (bombing by other means - killed)		
al yarmuk (9)	-0.31	HL	0.001
al amiriyah (16)	-0.31	HL	0.001
salim al hamadi (35)	-0.62	LH	0.005
al jaza' ir (36)	-0.62	LH	0.004
seha (40)	-0.62	LH	0.001
	627_6277IZ_6821IZ (bombing by other means - wounded)		
al amiriyah (16)	-0.33	HL	0.01
salim al hamadi (35)	-0.54	LH	0.01
al jaza' ir (36)	-0.54	LH	0.01
seha (40)	-0.54	LH	0.01
	627_6278IZ (rocket and or mortar attacks)		
al karradah (6)	-0.20	HL	0.001
ad dawrah (34)	-2.05	HL	0.001
al hurriyah (45)	-0.20	HL	0.001
	627_6278IZ_6820IZ (rocket and or mortar attacks - killed)		
al karradah (6)	-0.16	HL	0.001
al qadisiyah (11)	-0.03	HL	0.001
ad dawrah (34)	-1.87	HL	0.001
al hurriyah (45)	-0.16	HL	0.001
	627_6278IZ_6821IZ (rocket and or mortar attacks - wounded)		
al karradah (6)	-0.16	HL	0.001
al mansur (24)	-0.03	HL	0.001
ad dawrah (34)	-1.90	HL	0.001
al hurriyah (45)	-0.30	HL	0.001
	682 (offenses against life)		
al hurriyah (45)	-0.12	HL	0.001

	682_6820IZ (offenses against life - killed)		
al hurriyah (45)	-0.11	HL	0.001
	682_6820IZ_5631IZ (offenses against life - killed - sunni)		
al amiriyah (16)	-0.98	HL	0.001
ash shulah (23)	-0.43	LH	0.009
al mansur (24)	-0.98	HL	0.001
saydia dhubat (43)	-0.98	HL	0.001
al hurriyah (45)	-0.98	HL	0.001
	682_6821IZ (offenses against life - wounded)		
al amiriyah (16)	-0.20	HL	0.001
al hurriyah (45)	-0.20	HL	0.001
	683 (offenses against the person)		
ad dawrah (34)	-1.90	HL	0.001
	683_6830IZ (kidnapping indigenous person)		
ad dawrah (34)	-1.90	HL	0.001
	727 (aftermath of combat)		
al karkh (3)	-0.02	HL	0.001
ad dawrah (34)	-2.30	HL	0.001

	727_7270IZ (killed while captive)		
al karkh (3)	-0.02	HL	0.001
ad dawrah (34)	-2.30	HL	0.001
	727_7270IZ_5630IZ (collateral civilian wounded -sunni)		
al azamiyah (5)	-0.35	HL	0.001
al yarmuk (9)	-0.35	HL	0.001
ad dawrah (34)	-1.70	HL	0.001
al hurriyah (45)	-0.35	HL	0.001

dawrah) and #45 (al hurriyah). We also observe several instances where the locations identified for the encompassing variables (within city migration, hayy to hayy (166), social control (626), offenses against life (682), offenses against the person (683), and aftermath of combat (727)) show considerable overlap with more detailed subcategories.

Overall, these results need to be interpreted with care. The computation of the local Moran treats the variables as if they were measured on a ratio scale, which is clearly not the case here. Also, the presence of many zeros may lead to spurious results. We therefore proceed with an analysis that does not assume a ratio scale.

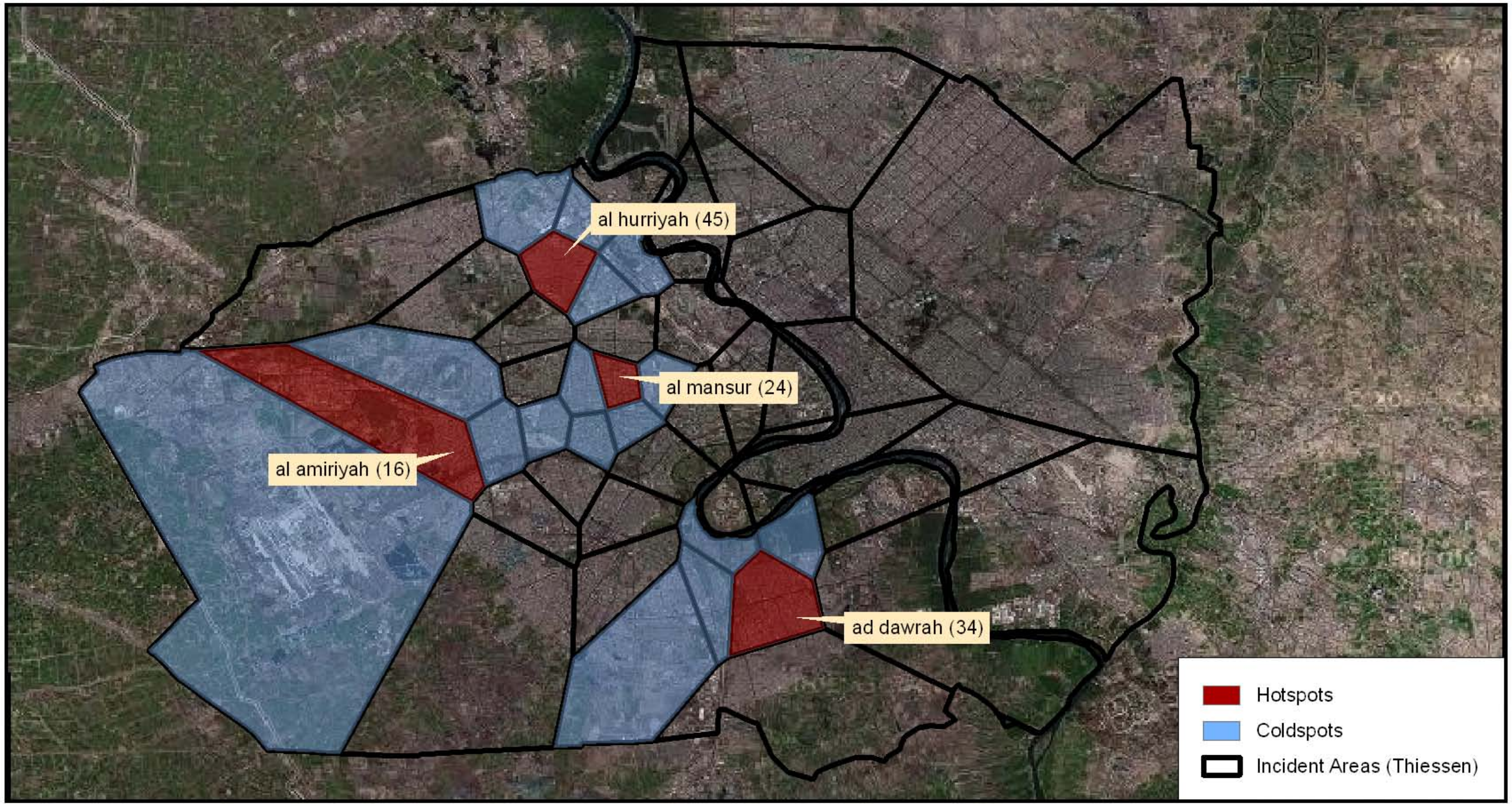


Figure 11: Four Neighborhoods (Spatial Outliers) with Largest Number of Significant Indicators (LISA Analysis)

al amiriyah (16):	**ad dawrah (34):**	**al hurriyah (45):**	**al mansur (24):**
within city migration, hayy to hayy	within city migration, hayy to hayy	social control	written threat
within city outmigration, hayy to hayy	within city outmigration, hayy to hayy	written threat	firing on crowd
social control	social control	car bombing	attack on political officials - killed
written threat	written threat	car bombing - killed	attack on mosque
car bombing	roadside or on road bombing - wounded	car bombing - wounded	attack on mosque - killed
car bombing - killed	firing on crowd	firing on crowd	rocket and or mortar attacks - wounded
roadside or on road bombing	attack on political officials	attack on mosque	offenses against life - killed - sunni
roadside or on road bombing - killed	attack on political officials - killed	attack on mosque - killed	
roadside or on road bombing - wounded	attack on mosque - killed	bombing by other means	
attack on political officials	rocket and or mortar attacks	rocket and or mortar attacks	
attack on political officials - killed	rocket and or mortar attacks - killed	rocket and or mortar attacks - killed	
attack on mosque	rocket and or mortar attacks - wounded	rocket and or mortar attacks - wounded	
bombing by other means	offenses against the person	offenses against life	
bombing by other means - killed	kidnapping indigenous person	offenses against life - killed	
bombing by other means - wounded	aftermath of combat	offenses against life - killed - sunni	
offenses against life - killed - sunni	killed while captive	offenses against life - wounded	
offenses against life - wounded	collateral civilian wounded -sunni	collateral civilian wounded -sunni	

Table 6: List of Significant Indicators for Four Neighborhoods (LISA)

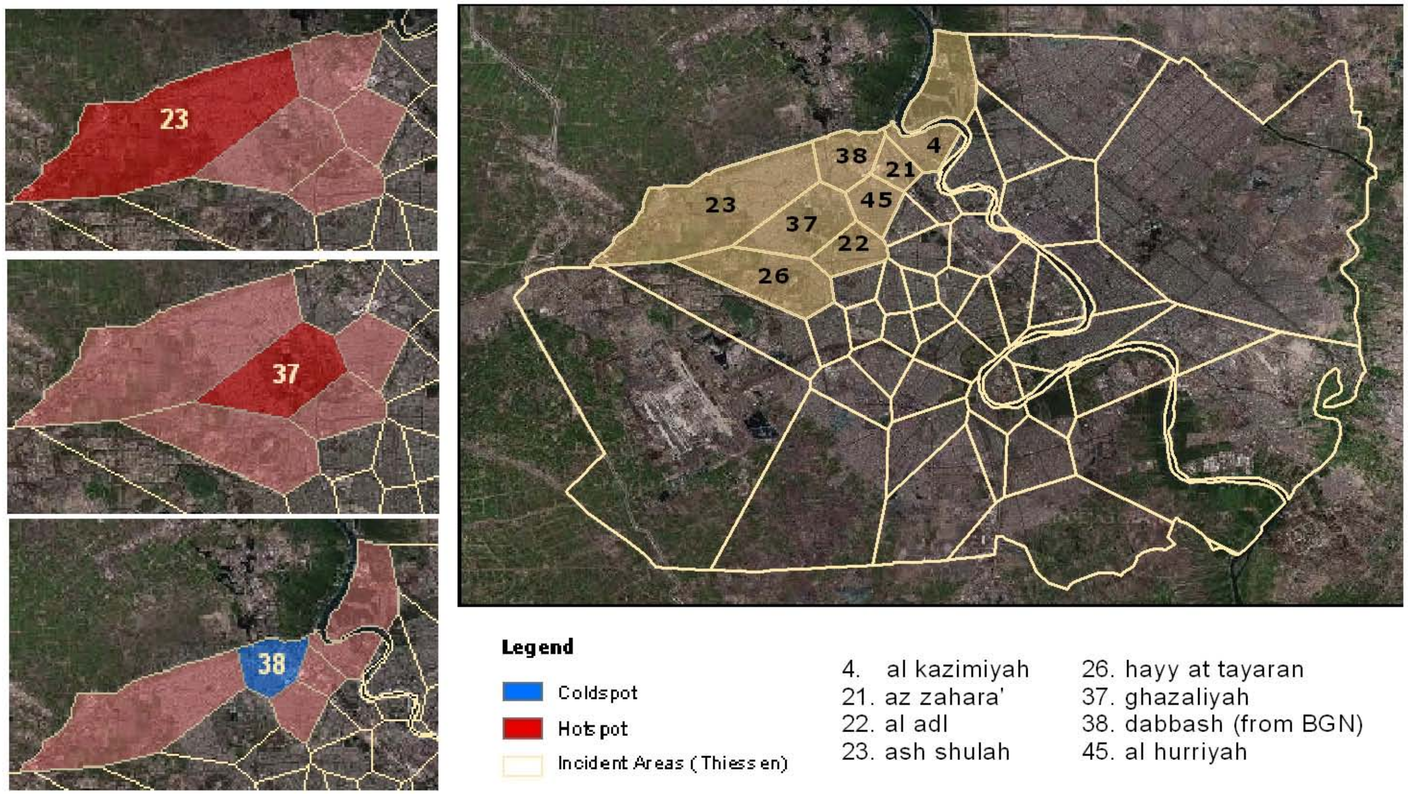


Figure 12: LISA Results for Within City In-Migration Hayy to Hayy

5.2 Binary Events

In this second part of the spatial analysis, the counts of events are transformed into binary 0-1 variables. While this involves a loss of information, it may shed light on different aspects of the spatial structure of the violent events. Also, given the source of the data (a content analysis of news sources) the actual count of events measured may be imprecise and too much importance may be given to the magnitude as such. Instead, we focus here on the presence-absence of violent events in the 45 neighborhoods.

As before, we start with an investigation of global patterns of spatial autocorrelation. We also explore local patterns to a limited extent. As discussed in Section 3.3.1, there remain some methodological issues to be investigated in this regard, so the results should be considered preliminary.

5.2.1 Global Spatial Autocorrelation

We assess global spatial autocorrelation by means of the *BB* join count statistic. We thus limit the focus to positive spatial autocorrelation between occurrences of violent events in neighborhoods. The WW statistic would measure positive spatial autocorrelation between the absence of events, which is less relevant. Given the sparsity of events, it is likely that the sample will be dominated by *BW* joins, so that a global pattern of negative spatial autocorrelation is to be expected. Hence, we do not further pursue this.

Table 7 reports the results for the *BB* join count statistics for the same 32 variables considered before. We use a permutation approach to assess significance.[5]

[5]Note that the *p*-value reported are pseudo significance levels. With 999 permutations, the highest level of significance that can be obtained is 0.001. This is a limitation of the permutation

Table 7: Join Count Statistics. Permutation Inference.

	BB	$E(BB)$	Var(BB)	p-value
166_1660IZ (within city inmigration, hayy to hayy)	1.250	0.221	0.036	0.001
166 (within city migration, hayy to hayy)	1.250	0.472	0.075	0.011
627_6278IZ_6821IZ (rocket and or mortar attacks - wounded)	1.750	1.016	0.133	0.034
627_6278IZ_6820IZ (rocket and or mortar attacks - killed)	1.750	1.030	0.157	0.051
682_6821IZ (offenses against life - wounded)	1.500	1.017	0.149	0.107
627_6277IZ_6821IZ (bombing by other means - wounded)	0.500	0.230	0.039	0.110
683_6830IZ (kidnapping indigenous person)	1.000	0.641	0.093	0.126
627_6277IZ (bombing by other means)	1.000	0.629	0.100	0.127
683 (offenses against the person)	1.000	0.639	0.099	0.128
682_6820IZ (offenses against life - killed)	2.875	2.357	0.266	0.155
626 (social control)	0.750	0.484	0.081	0.157
682 (offenses against life)	2.875	2.401	0.270	0.182
627_6276IZ (attack on mosque)	1.125	0.816	0.121	0.185
166_1661IZ (within city outmigration, hayy to hayy)	0.375	0.232	0.039	0.203
627_6277IZ_6820IZ (bombing by other means - killed)	0.500	0.335	0.052	0.221
627_6278IZ (rocket and or mortar attacks)	2.750	2.383	0.287	0.237
627_6276IZ_6820IZ (attack on mosque - killed)	0.500	0.344	0.055	0.238
626_6260IZ (written threat)	0.500	0.341	0.058	0.238
627_6271IZ (roadside or on road bombing)	1.000	0.804	0.104	0.256
627_6270IZ_6821IZ (car bombing - wounded)	1.000	0.821	0.121	0.272
627_6271IZ_6821IZ (roadside or on road bombing - wounded)	0.625	0.483	0.079	0.276
627_6275IZ (attack on political officials)	0.625	0.486	0.074	0.281
727 (aftermath of combat)	2.625	2.374	0.278	0.305
727_7270IZ (killed while captive)	2.625	2.398	0.295	0.326
627_6270IZ (car bombing)	1.375	1.246	0.156	0.361
727_7270IZ_5630IZ (collateral civilian wounded -sunni)	0.375	0.346	0.055	0.412
627_6273IZ (firing on crowd)	0.250	0.227	0.038	0.419
627_6270IZ_6820IZ (car bombing - killed)	1.000	1.015	0.140	0.467
627_6271IZ_6820IZ (roadside or on road bombing - killed)	0.625	0.654	0.101	0.505
627 (informal in group justice)	4.500	4.766	0.414	0.636
682_6820IZ_5631IZ (offenses against life - killed - sunni)	0.125	0.228	0.043	0.640
627_6275IZ_6820IZ (attack on political officials - killed)	0.000	0.237	0.039	0.885

The table contains four columns. The statistic itself is listed in the first column, the mean and variance obtained from the permuted samples in columns two and three, and the p-value in column four.

As in Section 5.1.1, there is again strong evidence for positive and significant spatial autocorrelation of the variable 166_1660IZ (within city inmigration, hayy to hayy) at $p = 0.001$. In addition, the encompassing variable 166 (internal migration) also shows significant positive spatial autocorrelation, at $p = 0.011$. The variables 627_6277IZ_6820IZ (bombing by other means – killed) and 627_6277IZ_6821IZ (bombing by other means – wounded) are no longer significant. Instead, two new variables appear, 627_6278IZ_6821IZ (rocket and mortar attacks – wounded) at $p = 0.034$, and marginally significant 627_6278IZ_6820IZ (rocket and mortar attacks – killed) at $p = 0.051$.

In sum, there is only limited support for significant global spatial pattern in the violent events across neighborhoods.

5.2.2 Local Spatial Autocorrelation

In this final section, we begin to address patterns of local spatial autocorrelation for binary variables. As pointed out in Section 3.3.1, there are two important concepts in this regard, composition and configuration. Since the data set consists of only 45 observations, it was not possible to address configuration. For example, in Boots (2003), the method for quantifying configuration relies on a moving window of up to 7×7, which would be meaningless in the current situation. Even the smallest window, or 3×3 would take up one fifth of the data set for each location, which is clearly inappropriate.

method.

Table 8: Local Composition

	Significant Locations
166(within city migration, hayy to hayy)	al adl (22) ash shulah (23) ghazaliyah (37)
166_1660IZ (within city inmigration, hayy to hayy)	al adl (22) ash shulah (23) ghazaliyah (37)
626 (social control)	ash shulah (23) ghazaliyah (37)
626_6260IZ (written threat)	ash shulah (23) ghazaliyah (37)
627_6271IZ_6820IZ (roadside or on road bombing - killed)	al ma'mun (10)
627_6271IZ_6821IZ (roadside or on road bombing - wounded)	al ma'mun (10)
627_6275IZ (attack on political officials)	al ma'mun (10)
627_6276IZ_6820IZ (attack on political officials - killed)	al kazimiyah (4) dabbash (38)
627_6277IZ (bombing by other means)	ash shulah (23) ghazaliyah (37)
627_6278IZ (rocket and or mortar attacks)	al quds (2) al qahirah (7) ash shulah (23) ghazaliyah (37)
627_6278IZ_6820IZ (rocket and or mortar attacks - killed)	ash shulah (23) ghazaliyah (37)
627_6278IZ_6821IZ (rocket and or mortar attacks - wounded)	ash shulah (23) ghazaliyah (37)
682 (offenses against life)	al adl (22) ash shulah (23) ghazaliyah (37)
682_6820IZ (offenses against life - killed)	al adl (22) ash shulah (23) ghazaliyah (37)
682_6821IZ (offenses against life - wounded)	al adl (22)
683 (offenses against the person)	al ma'mun (10) ash shulah (23) ghazaliyah (37)
683_6830IZ (kidnapping indigenous person)	al ma'mun (10) ash shulah (23) ghazaliyah (37)
727 (aftermath of combat)	al adl (22) ash shulah (23) ghazaliyah (37)
727_7270IZ (killed while captive)	al adl (22) ash shulah (23) ghazaliyah (37)
727_7270IZ_5630IZ (collateral civilian wounded -sunni)	ash shulah (23) ghazaliyah (37)

We assess composition by computing the probability of observing the number of neighborhoods with an event ($x = 1$), in a window defined by the four nearest neighbors. This probability can be computed analytically as in Equation (13). Those locations for which the result was less than $p = 0.05$ are listed in Table 8. Only variables that included at least one significant location are included.

Of the 45 variables, 20 show significant local composition. However, the number of locations for each is rather small, ranging from one for four of the variables, to four for 627_6278IZ (rocket and mortar attacks). Compared to the Local Moran for the full counts, there are both fewer variables and fewer locations per variable. Again, we observe several instances where the encompassing variable (within city migration, hayy to hayy (166), social control (626), offenses against life (682), and offenses against the person (683)) shows the same pattern as more detailed subcategories.

Interestingly, a small number of locations appear as local clusters for several of the variables (Figure 13 and Table 9). Specifically, #23 (ash shulah) and #37 (ghazaliyah) appear no less than 15 times, #22 (al adl) is listed seven times and #10 (al mamun) five times. Note that these are not the same neighborhoods as the ones identified by means of the Local Moran. Also, both # 23 and #37 are boundary locations, which may suffer from spurious inference due to the boundary effect. In contrast, #10 is a central location, which deserves further scrutiny.

These findings, although preliminary, would suggest that the major locus of violent events tends to be centered around a few neighborhoods (the ones identified together with their four nearest neighbors). However, this only addresses the composition aspect of local clustering. Further methodological refinement is needed to extend the regular grid case to irregular spatial configurations. Also,

the difference between the neighborhoods identified by means of spatial autocorrelation measures that use the full count and the ones based on binary variables remains to be further investigated.

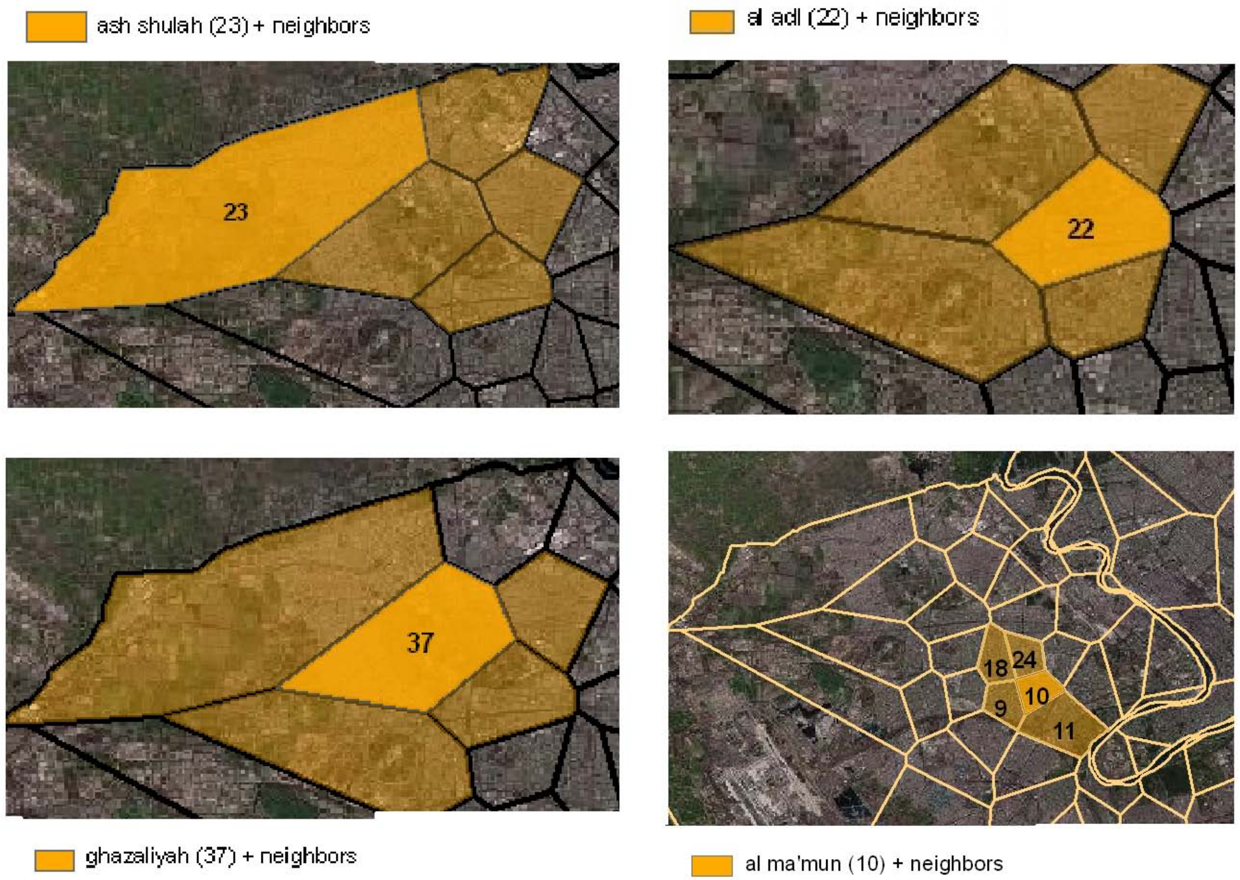


Figure 13: 4 Neighborhoods with Largest No. of Sign. Indicators (Join Count)

ash shulah (23):	ghazaliyah (37):	al adl (22):	al ma'mun (10):
within city migration, hayy to hayy	within city migration, hayy to hayy	within city migration, hayy to hayy	car bombing - killed
within city inmigration, hayy to hayy	within city inmigration, hayy to hayy	within city inmigration, hayy to hayy	car bombing - wounded
social control	social control	offenses against life	attack on political officials
written threat	written threat	offenses against life - killed	offenses against the person
bombing by other means	bombing by other means	offenses against life - wounded	kidnapping indigenous person
rocket and or mortar attacks	rocket and or mortar attacks	aftermath of combat	
rocket and or mortar attacks - killed	rocket and or mortar attacks - killed	killed while captive	
rocket and or mortar attacks - wounded	rocket and or mortar attacks - wounded		
offenses against life	offenses against life		
offenses against life - killed	offenses against life - killed		
offenses against the person	offenses against the person		
kidnapping indigenous person	kidnapping indigenous person		
aftermath of combat	aftermath of combat		
killed while captive	killed while captive		
collateral civilian wounded -sunni	collateral civilian wounded -sunni		

Table 9: List of Significant Indicators for Four Neighborhoods (Join Count)

6 Conclusions

This preliminary analysis of the spatial pattern of violent events in Baghdad only allows for limited conclusions. The sparseness of events over the time period considered and at the spatial scale of a neighborhood may not yield sufficient information to fully assess the characteristics of their spatial distribution across the 45 sites.

For 80 of the variables contained in the original classification that was constructed from the text sources (news reports), there were insufficient observations to allow for any statistical analysis. Of the remaining 32 variables, a few show consistent spatial patterns, both at the global and local scale. The only variable where this evidence is strong is 166_1660IZ (within city inmigration, hayy to hayy), which consistently shows signs of positive spatial autocorrelation, or clustering (Figure 12). In contrast, the evidence for the other variables tends to point to negative spatial autocorrelation, suggesting that neighborhoods with violent events tend to be surrounded by neighborhoods without those events (Figure 11).

Interestingly, especially for the measures constructed for binary variables, a handful of neighborhoods seem to show local spatial patterns for a range of indicators of violent events (Figure 11 and Figure 13). This may warrant further attention, by complementing the current quantitative analysis with on the ground ethnographic information.

Greater precision in georeferencing may provide the detail needed to carry out an analysis at a spatial scale smaller than the current neighborhood. In addition, the consideration of different time periods may provide larger samples of violent events. Finally, some methodological refinement is needed to extend the join

count statistics to the local context in a satisfactory way.

References

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2):93–115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In Fisher, M., Scholten, H., and Unwin, D., editors, *Spatial Analytical Perspectives on GIS*, pages 111–125. London, Taylor and Francis.

Anselin, L. and Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A. and Giles, D., editors, *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, New York.

Anselin, L., Syabri, I., and Kho, Y. (2006). Geoda: An introduction to spatial data analysis. *Geographical Analysis*, 38(1):5–22.

Bivand, R. (2001). More on spatial data analysis. *R News*, 1(3):13–17.

Bivand, R. (2002). Spatial econometrics functions in r: classes and methods. *Journal of Geographical System*, 4:405–421.

Bivand, R. (2006). Implementing spatial data analysis software tools in r. *Geographical Analysis*, 38(1):23–40.

Bivand, R. S. and Gebhardt, A. (2000). Implementing functions for spatial statistical analysis using the r language. *Journal of Geographical System*, 2:307–312.

Bivand, R. S. and Portnov, B. A. (2004). Exploring spatial data analysis techniques using r: The case of observations with no neighbors. In Anselin, L., Florax, R. J., and Rey, S. J., editors, *Advances in Spatial Econometrics: Methodology, Tools and Applications*, page 121142. Springer-Verlag, Berlin.

Boots, B. (2003). Developing local measures of spatial association for categorical data. *Journal of Geographical System*, 5:139–160.

Case, A. (1991). Spatial patterns in household demand. *Econometrica*, 59:953–965.

Case, A. (1993). Neighborhood influence and technological change. *Regional Science and Urban Economics*, 22:491–508.

Case, A., Rosen, H., and Hines, J. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52:285–307.

Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*. Pion, London.

Cliff, A. and Ord, J. (1981). *Spatial Processes: Models and Applications*. Pion, London.

Conley, T. and Ligon, E. (2002). Economic distance, spillovers and cross country comparison. *Journal of Economic Growth*, 7:157–187.

Conley, T. and Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17:303–327.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York.

Geary, R. (1954). The contiguity ratio and statistical mapping. *The incorporated Statistician*, 5:115–145.

Getis, A. and Ord, J. (1996). Local spatial statistics: an overview. In Longley, P. and Batty, M., editors, *Spatial Analysis: modelling in a GIS environment*, pages 261–277. GeoInformation International, Cambridge.

Li, H. and Reynolds, J. (1993). A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecology*, 8:155–162.

Li, H. and Reynolds, J. (1994). A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology*, 75:2446–2455.

Li, H. and Reynolds, J. (1995). On definition and quantification of heterogeneity. *Oikos*, 73:280–284.

Moran, P. (1950). Note on continuous stochastic phenomena. *Biometrika*, 37:17–23.

Pinkse, J., Slade, M. E., and Brett, C. (2002). Spatial price competition: A semiparametric approach. *Econometrica*, 70(3):1111–1153.

Schabenberger, O. and Gotway, C. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, Boca Raton, FL.

Sokal, R., N.Oden, and Thomson, B. (1998). Local spatial autocorrelation in a biological model. *Geographical Analysis*, 30:331–354.

Appendix A: List of Neighborhoods

1. al mastansiriyah
2. al quds
3. al karkh
4. al kazimiyah
5. al azamiyah
6. al karradah
7. al qahirah
8. habibiya
9. al yarmuk
10. al ma'mun
11. al qadisiyah
12. al jahid
13. abd al wahid
14. al habra
15. hayy al furat
16. al amiriyah
17. al firdaws
18. al andalus
19. al kindi
20. zidan
21. az zahara'
22. al adl
23. ash shulah

24. al mansur
25. al khadra
26. hayy at tayaran
27. as salam
28. qaryat washash
29. al muradiyah
30. jabbur
31. at ta'mim
32. al marifah
33. basatin jamil wadi
34. ad dawrah
35. salim al hamadi
36. al jaza'ir
37. ghazaliyah
38. dabbash (from BGN)
39. shurtah
40. seha
41. zawra park
42. old al muthanna airport
43. saydia dhubat
44. daoudi
45. al hurriyah

Appendix B: Classification of Events

- 103_Place_Name ad_dawrah
- 166_Internal_migration_Hayy_to_Hayy
  - 1661IZ_Out
    - 5630IZ_shia
    - 5631IZ_sunni
    - 5632IZ_christian
    - 5633IZ_kurd
    - 5634IZ_chaldean
  - 1660IZ_IN
    - 5630IZ_shia
    - 5631IZ_sunni
    - 5632IZ_christian
    - 5633IZ_kurd
    - 5634IZ_chaldean
- 167_external_migration
  - 1671IZ_Into_Baghdad
    - 5630IZ_shia
    - 5631IZ_sunni
    - 5632IZ_christian
    - 5633IZ_kurd
    - 5634IZ_chaldean
  - 1670IZ_Out_of_Baghdad
    - 5630IZ_shia
    - 5631IZ_sunni
    - 5632IZ_christian
    - 5633IZ_kurd
    - 5634IZ_chaldean
- 626_social_control
  - 6260IZ_written_threat

    * 5631IZ_sunni
    * 5630IZ_shia

- 627_Informal_in_group_justice
  - 6270IZ_Car_bombing
    * 7278IZ_collateral_civilian_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7279IZ_collateral_civilian_dead
      · 5631IZ_sunni
      · 5630IZ_shia
    * 6821IZ_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 6820IZ_killed
      · 5631IZ_sunni
      · 5630IZ_shia
  - 6271IZ_roadside_or_on_road_bombing
    * 6821IZ_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 6820IZ_killed
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7278IZ_collateral_civilian_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7279IZ_collateral_civilian_dead
      · 5631IZ_sunni
      · 5630IZ_shia
  - 6273IZ_firing_on_crowd
    * 6820IZ_killed
      · 5630IZ_shia
      · 5631IZ_sunni
    * 6821IZ_wounded

        - 5630IZ_shia
        - 5631IZ_sunni
- 6275IZ_attack_on_political_official
    - 6821IZ_wounded
        - 5631IZ_sunni
        - 5630IZ_shia
    - 6820IZ_killed
        - 5631IZ_sunni
        - 5630IZ_shia
- 6276IZ_attack_on_mosque
    - 6821IZ_wounded
        - 5631IZ_sunni
        - 5630IZ_shia
    - 6820IZ_killed
        - 5631IZ_sunni
        - 5630IZ_shia
- 6277IZ_bombing_by_other_means
    - 7278IZ_collateral_civilian_wounded
        - 5631IZ_sunni
        - 5630IZ_shia
    - 7279IZ_collateral_civilian_dead
        - 5631IZ_sunni
        - 5630IZ_shia
    - 6821IZ_wounded
        - 5631IZ_sunni
        - 5630IZ_shia
    - 6820IZ_killed
        - 5631IZ_sunni
        - 5630IZ_shia
- 6278IZ_rocket_and_or_mortar_attacks
    - 7278IZ_collateral_civilian_wounded
        - 5631IZ_sunni
        - 5630IZ_shia
    - 7279IZ_collateral_civilian_dead

      - 5631IZ_sunni
      - 5630IZ_shia
    - 6821IZ_wounded
      - 5631IZ_sunni
      - 5630IZ_shia
    - 6820IZ_killed
      - 5631IZ_sunni
      - 5630IZ_shia

- 682_offenses_against_life
  - 6821IZ_wounded
    - 5631IZ_sunni
      - 463?IZ_insert_occupation_code
    - 5630IZ_shia
      - 463?IZ_insert_occupation_code
  - 6820IZ_killed
    - 5631IZ_sunni
      - 4651IZ_electrician
    - 5630IZ_shia
      - 4671IZ_police_officer
- 683_offenses_against_the_person
  - 6830IZ_Kidnapping_indigenous_person
    - 5631IZ_sunni
      - 4666IZ_minister
    - 5630IZ_shia
      - 463?IZ_doctor_physician
- 727_Aftermath_of_combat
  - 7270IZ_killed_while_captive
    - 5631IZ_sunni
    - 5630IZ_shia
  - 7278IZ_collateral_civilian_wounded
    - 5631IZ_sunni

    * 5630IZ_shia
  - 7279IZ_collateral_civilian_dead
    * 5631IZ_sunni
    * 5630IZ_shia
- 762_Suicide
  - 7620IZ_suicide_bombing
    * 6821IZ_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 6820IZ_killed
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7278IZ_collateral_civilian_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7279IZ_collateral_civilian_dead
      · 5631IZ_sunni
      · 5630IZ_shia
  - 7622IZ_suicide_bombing_vehicular_born
    * 6821IZ_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 6820IZ_killed
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7278IZ_collateral_civilian_wounded
      · 5631IZ_sunni
      · 5630IZ_shia
    * 7279IZ_collateral_civilian_dead
      · 5631IZ_sunni
      · 5630IZ_shia
  - 7621IZ_suicide_bombing_individual_born
    * 6821IZ_wounded

      - 5631IZ_sunni
      - 5630IZ_shia
    - 6820IZ_killed
      - 5631IZ_sunni
      - 5630IZ_shia
    - 7278IZ_collateral_civilian_wounded
      - 5631IZ_sunni
      - 5630IZ_shia
    - 7279IZ_collateral_civilian_dead
      - 5631IZ_sunni
      - 5630IZ_shia

- 788_Rituals
  - 7881IZ_Pilgrimages
    - 5360IZ_shia
      - 6821IZ_wounded
      - 6820IZ_killed
  - 7880IZ_Formal_processions
    - 5360IZ_shia
      - 6821IZ_wounded
      - 6820IZ_killed
    - 5361IZ_sunni
      - 6821IZ_wounded
      - 6820IZ_killed

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)
17-05-2009

2. REPORT TYPE
Final

3. DATES COVERED (From - To)

4. TITLE AND SUBTITLE
Approaches Towards the Identification of Patterns in Violent Events, Baghdad, Iraq

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT

6. AUTHOR(S)
Luc Anselin and Gianfranco Piras

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
PO Box 9005,
Champaign, IL 61826-9005

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CERL CR-09-1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Assistant Secretary of the Army for Acquisition, Logistics, and Technology (ASAALT)
2511 Jefferson Davis Highway, Presidential Towers
ATTN: FOIA OFFICER, RM 11500
Arlington, VA 22202-3911

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This work analyzed the spatial distribution of violent events as constructed from a content analysis of open source news reports. Data on 112 variables was available for 45 neighborhoods. The small sample limited the analysis to those 32 variables with at least four observations. The statistical analysis was done both for the original measures of event counts by neighborhood, and for binary variables that indicated the presence of events. Test statistics for spatial autocorrelation were computed for global patterns and local patterns, including global and local Moran's I, Geary's c, Moran Scatterplot, join count statistics, and local join counts. There was little evidence of systematic spatial structure at the neighborhood scale. Only for a variable indicating internal between hayy migration was there consistent indication of positive spatial autocorrelation, or clustering. Several other variables showed significant negative spatial autocorrelation at the local scale, suggesting that neighborhoods where violent events occurred are surrounded by neighborhoods without violent events. A few neighborhoods were consistently identified as the locus of a spatial outlier, suggesting some patterning. A finer spatial scale might reveal more complex spatial patterns. The current data do not allow this to be investigated.

15. SUBJECT TERMS
military, LISA, join count, content analysis, qualitative, spatial analysis

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER (include area code)
Unclassified	Unclassified	Unclassified	SAR	84	